

# The Degrees of Freedom Region and Interference Alignment for the MIMO Interference Channel with Delayed CSI

Chinmay S. Vaze and Mahesh K. Varanasi

## Abstract

The degrees of freedom (DoF) region of the 2-user multiple-antenna or MIMO (multiple-input, multiple-output) interference channel (IC) is studied under fast fading and the assumption of *delayed* channel state information (CSI) wherein all terminals know all (or certain) channel matrices perfectly, but with a delay, and each receiver in addition knows its own incoming channels instantaneously. The general MIMO IC is considered with an arbitrary number of antennas at each of the four terminals. Dividing it into several classes depending on the relation between the numbers of antennas at the four terminals, the fundamental DoF regions are characterized under the delayed CSI assumption for *all* possible values of number of antennas at the four terminals. In particular, an outer bound on the DoF region of the general MIMO IC is derived. This bound is then shown to be tight for all MIMO ICs by developing interference alignment based achievability schemes for each class. A comparison of these DoF regions under the delayed CSI assumption is made with those of the idealistic ‘perfect CSI’ assumption where perfect and instantaneous CSI is available at all terminals on the one hand and with the DoF regions of the conservative ‘no CSI’ assumption on the other, where CSI is available at the receivers but not at all at the transmitters.

## Index Terms

Degrees of freedom, Delayed CSI, Fading channels, Interference alignment, Interference channel, MIMO, Outer bound.

This work was supported in part by NSF Grant CCF-0728955. The authors are with the Department of Electrical, Computer, and Energy Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: Chinmay.Vaze, varanasi@colorado.edu).

## I. INTRODUCTION

**T**HE interference channel (IC) is an archetype of a network for multiple unicast communication. This channel consists of two transmit-receive pairs that wish to communicate over a common noisy medium. Despite over three decades of research the capacity region of the IC remains unknown in general. Much progress has been made towards this goal however, in the form of capacity characterizations of special classes of the interference channel [1]–[8] and capacity approximations for the scalar (single-antenna) Gaussian IC [9] and the vector or MIMO Gaussian IC [10], [11], where the capacity region is characterized to within a constant gap that is independent of signal-to-noise ratio (SNR) and the channel parameters. These works assume perfect channel state information (CSI) at all terminals.

For the MIMO Gaussian IC, under the same perfect CSI assumption, the degrees of freedom (DoF) region – which denotes the rate of growth with respect to  $\log \text{SNR}$  of the capacity region at high SNR – was found in [12], [13]. As shown there, the DoF region can be achieved via a combination of transmit/receive zero-forcing beamforming with time sharing, wherein the transmitters exploit the channel null spaces to beamform transmitted signals so as to cause minimal interference at the unintended receiver. However, such schemes can be sensitive to imperfections in CSI at the transmitters (CSIT). Indeed, it has been recently proved that in the extreme case of having no CSI whatsoever at the transmitters, the DoF region of the channel collapses to the extent that just receive zero-forcing beamforming (and time sharing) are DoF-region optimal [14]–[16].

The impact of lack of CSIT on the achievable rate emphasizes the need for transmitters to obtain CSI from the receivers. The receivers can learn their current incoming channel state via pilot transmissions with a high degree of accuracy<sup>1</sup>, which in practice, can be broadcast back to the other terminals via a feedback link. As a result, CSI available at different terminals of the network may be neither perfect, due to the limited capacity of the feedback link, nor instantaneous, due to the delays involved in channel estimation and feedback. The first problem of having imperfect CSI has been relatively well investigated and it is now known in the variety of contexts that if the quality of CSI available at the terminals can be improved at a sufficiently fast rate with  $\log \text{SNR}$ , then the perfect CSI DoF can be achieved even with imperfect CSIT (cf.

<sup>1</sup>Throughout this paper, it is assumed that receivers know their incoming channel state perfectly and instantaneously.

[17]–[19], and references therein). However, fundamental approaches to the study of the second problem of having just delayed CSI, a setting that is highly relevant in mobile environments with short channel coherence times, has only recently begun [20]–[22].

In particular, Maddah-Ali and Tse in [20] prove in the context of the MISO broadcast channel (BC) (ie., a BC with multiple-antenna transmitter and single antenna receivers) the surprising and interesting result that the presence of delayed CSI at the terminals can lead to a dramatic increase in the achievable rate relative to having no CSI to the extent that even the DoF of the channel increase. Moreover, this is shown to be possible even if the channel coefficients vary independently across each time slot so that the delayed CSI available to the terminals is actually outdated (or ‘stale’). As a case in point, in the MISO BC with 2 transmit antennas, 2 single-antenna users, and i.i.d. fast fading, the achievable sum-DoF improve from 1 with no CSI to  $\frac{4}{3}$  with delayed CSI. The achievable scheme in [20] is based on the idea that the interference encountered at a prior time instant by one of the users, which is a linear combination (LC) of the data symbols intended for some other user, can be perfectly evaluated by the transmitter at the current time instant using delayed CSI, and hence, can be subsequently transmitted to provide an opportunity to the interfered user to learn that interference while simultaneously providing a new LC to the user where these symbols are desired. Furthermore, the achievability scheme based on this idea has been shown there to be sum-DoF optimal for the class of MISO BCs in which there are at least as many transmit antennas as the number of users. This work was recently extended by the authors of this paper to the case of the MIMO BC (with multiple and arbitrary number of antennas at each terminal) in [22] where an outer-bound to the DoF region is obtained for the general  $K$ -user case, and was proved, through an interference alignment based achievability scheme, to be tight in the 2-user case.

Meanwhile, Maleki et al in [21] show how the so-called blind interference alignment scheme for the  $(3, 1, 4, 2)$  MIMO IC (i.e., with 3 and 1 antennas at transmitters one and two, and 4 and 2 antennas at the receivers one and two, respectively) obtained earlier for the staggered block fading model in [23] can be adapted to the same MIMO IC but under the delayed CSI model to achieve the  $(1, 1.5)$  DoF tuple, a point that lies outside the DoF region with no CSI [15]. However, with no outer bound for that example, the question of whether this example can be improved is left open and so is the much more general question of characterizing the DoF region of the  $(M_1, M_2, N_1, N_2)$  MIMO IC (with  $M_i$  antennas at transmitter  $i$  and  $N_i$  antennas

at receiver  $i$  for  $i \in \{1, 2\}$ ). Two more schemes were also proposed in [21] under the delayed CSI model for the 2-user scalar X channel (with a sum DoF of  $8/7$ ) and the 3-user scalar IC (with a sum DoF of  $9/8$ ) but the question of DoF-optimality in these cases also remains open as is the characterization of the DoF regions.

In this paper, we study the problem of characterizing the DoF region with delayed CSI of the general  $(M_1, M_2, N_1, N_2)$  MIMO IC. In particular, we first obtain an outer-bound for the general case. Toward this end, Fano's inequality is used to upper-bound the rate achievable for each user by the mutual information between the signal received by that user and its desired message, and then a technique is developed to bound the DoF of a certain key weighted sum of these two mutual information terms to finally obtain an upper-bound on that weighted sum of the DoF achievable for the two users (see Sections IV and V). Note that apart from the 2-user MIMO IC considered here, the delayed-CSI DoF region is known only for the MISO BC [20] and its generalization to MIMO BC in [22], which prompts us to contrast the techniques developed in this work to that in [20], [22]. In order to derive their outer bound, the authors in [20] first outer bound the capacity region by transforming the original BC through genie-aided side information into a physically-degraded BC with a transmitter that knows the past channel states and past channel outputs and then compute the DoF region of the resultant physically-degraded BC by using the result of [24] that the capacity of the physically-degraded BC with feedback is the same as that without feedback. However, the reliance on the result of [24] makes the technique in [20] not applicable to other channels with delayed CSI, such as the IC. In contrast, the general idea developed here can be used for dealing with a variety of channels with delayed CSI. For instance, using our technique, it is possible to provide an alternate proof for the outer-bound of [20] and [22] for the MISO and MIMO BCs without invoking the result of [24]. Moreover, the same technique could also be employed with the MIMO cognitive IC under the delayed CSI assumption. The cognitive IC is an interference channel in which one or more terminals can be simultaneously cognitive (cf. [13], [25], [26] and references therein).

Next, the general  $(M_1, M_2, N_1, N_2)$  MIMO IC is divided into several classes of MIMO ICs depending on the relative numbers of antennas at the four terminals. Interference alignment based communication schemes specifically tailored for each class (when necessary) are given and shown to have DoF regions that coincide with the outer bound for all classes. Hence, the fundamental DOF region is obtained for the general  $(M_1, M_2, N_1, N_2)$  MIMO IC without any

restriction on the numbers of antennas. The MIMO IC, unlike the MISO or MIMO BC, consists of distributed, or non-cooperating, transmitters. Hence, neither transmitter can compute the past received signals (excluding noise terms) of the two receivers even with delayed CSI, unlike the case with the transmitter in a BC. However, since the transmit signal of any given transmitter is intended for just its paired receiver, each transmitter can compute the interference caused by its signal to the unintended receiver at previous time instants. Thus, interference alignment can be achieved by requiring the transmitter(s) to transmit the interference they created in the past, which provides additional information to the intended receiver about its desired message without causing any extra interference at their unpaired receiver. Interestingly, there is even a class of MIMO ICs – in which the no-CSI DoF region is strictly smaller than the perfect-CSI DoF region – for which the DoF region with delayed CSI is seen to coincide with the entire perfect-CSI DoF region.

Incidentally, as a result of this work, it can now be asserted that the  $(1, 1.5)$  DoF pair achievable using the retrospective interference alignment scheme of [21] over the  $(3, 1, 4, 2)$  MIMO IC lies on the boundary of the DoF region (see Lemma 12 in Section VIII which derives the DoF region for a class of MIMO ICs that contains the  $(3, 1, 4, 2)$  MIMO IC).

A complete comparative characterization of the DoF regions of the MIMO IC under the delayed CSI assumption is given with the DoF regions under the idealized assumption of perfect CSI [12], [13] on the one hand and with the DoF regions under the conservative assumption of no CSI [14]–[16] on the other, thereby revealing a rich classification of MIMO ICs according to whether the no CSI DoF region is strictly contained by (or is equal to) the delayed CSI DoF region which in turn is strictly contained by (or is equal to) the perfect CSI DoF region.

The rest of the paper is organized as follows. In Section II, the MIMO IC model is defined. Section III contains the main results, namely, the outer and inner bounds on the DoF region with delayed CSI. Sections IV–X provide the proofs of the main results. Section XI concludes the paper.

## II. THE CHANNEL MODEL

In this section, we describe the  $(M_1, M_1, N_1, N_2)$  MIMO IC with fading under the assumption of delayed CSI. The two transmitters of the MIMO IC are denoted as T1 and T2, with transmitter  $i$  having  $M_i$  antennas and their corresponding receivers are denoted as R1 and R2, with receiver

$i$  equipped with  $N_i$  antennas. A given transmitter has a message only for its respective/paired receiver. However, its signal is received at the unintended receiver as interference. The input-output relationship at the  $t^{th}$  channel use is given at R1 and R2, respectively, as

$$Y_1(t) = H_{11}(t)X_1(t) + H_{12}(t)X_2(t) + W_1(t), \quad (1)$$

$$Y_2(t) = H_{21}(t)X_1(t) + H_{22}(t)X_2(t) + W_2(t), \quad (2)$$

where  $X_i(t)$  is the signals transmitted by the  $i^{th}$  transmitter  $T_i$ ;  $H_{ij}(t) \in \mathbb{C}^{N_i \times M_j}$  denotes the channel matrix between  $i^{th}$  receiver and  $j^{th}$  transmitter; and  $W_i(t) \sim \mathcal{CN}(0, I_{N_i})$ , for  $i = 1, 2$ , is the additive noise at receiver  $i$ . The power is constrained to be  $P$ , i.e.,  $\mathbb{E}[\|X_i(t)\|^2] \leq P \forall i, t$ .

For simplicity, we assume the case of Rayleigh fading, i.e., all the entries of all channel matrices  $\{H_{ij}(t)\}_{i,j}$  are independent, identically distributed (i.i.d.) zero-mean, unit variance complex normal (denoted  $\mathcal{CN}(0, 1)$ ) random variables. Further, the channel and noise realizations are taken to be i.i.d. across time and they are independent of each other.

It is assumed that every receiver knows the channel matrices corresponding to itself perfectly and instantaneously, while all other terminals know them perfectly with a delay of one time unit. In particular, for each  $i \in \{1, 2\}$ , the channel matrices  $H_{i1}(t)$  and  $H_{i2}(t)$  are known to the  $i^{th}$  receiver at time  $t$ , whereas all other terminals (i.e.,  $T_1$ ,  $T_2$ , and the  $j^{th}$  receiver with  $j \neq i$ ) know these matrices at time  $t + 1$ . We refer to this assumption about CSI as the “delayed CSI” assumption.

The case where all terminals have perfect and instantaneous knowledge of all channel matrices is referred to as the “perfect CSI” assumption. If the  $i^{th}$  receiver knows the channel matrices  $H_{i1}(t)$  and  $H_{i2}(t)$  perfectly and instantaneously and if there is no additional CSI at any of the terminals, then we refer to the corresponding case as the “no CSI” assumption. It must be noted that in all the three cases, all terminals are always assumed to know the distribution of the channel matrices.

Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two independent messages to be sent by  $T_1$  and  $T_2$ , respectively, over a block length of  $n$ , where the message  $\mathcal{M}_i$  is intended for the  $i^{th}$  receiver. It is assumed that  $\mathcal{M}_i$  is distributed uniformly over a set of cardinality  $2^{nR_i(P)}$ , when there is a power constraint of  $P$  at the transmitters. If  $\overline{H}(n) = \{H_{11}(t), H_{12}(t), H_{21}(t), H_{22}(t)\}_{t=1}^n$  with  $\overline{H}(0) = \phi$  and  $\overline{Y}_i(n) = \{Y_i(t)\}_{t=1}^n$ , then a coding scheme for blocklength  $n$  consists of two encoding functions

$f_i^{(n)} = \{f_{i,t}^{(n)}\}_{t=1}^n$ ,  $i = 1, 2$ , such that

$$X_i(t) = f_{i,t}^{(n)}\left(\mathcal{M}_i, \overline{H}(t-1)\right) \quad \forall t \in \{1, 2, \dots, n\}$$

and two decoding functions such that

$$\hat{\mathcal{M}}_i = g_i^{(n)}\left(\overline{Y}_i(n), \overline{H}(n-1), H_{i1}(n), H_{i2}(n)\right) \text{ where } i \in \{1, 2\}.$$

A rate tuple  $(R_1(P), R_2(P))$  is said to be achievable if there exists a sequence of coding schemes such that probability of  $\mathcal{M}_1 \neq \hat{\mathcal{M}}_1$  or  $\mathcal{M}_2 \neq \hat{\mathcal{M}}_2$  tends to zero as  $n \rightarrow \infty$ . The capacity region  $\mathcal{C}(P)$  is defined to be the set of all achievable rate tuples  $(R_1(P), R_2(P))$  when the power constraint at the transmitters is  $P$ . We then define the DoF region as

$$\mathbf{D}^{\text{d-CSI}} = \left\{ (d_1, d_2) \left| d_i \geq 0 \text{ and } \exists (R_1(P), R_2(P)) \in \mathcal{C}(P) \text{ s.t. } d_i = \lim_{P \rightarrow \infty} \frac{R_i(P)}{\log_2 P} \text{ } i \in \{1, 2\} \right. \right\}.$$

The DoF regions corresponding to the cases of perfect and no CSI, denoted respectively as  $\mathbf{D}^{\text{p-CSI}}$  and  $\mathbf{D}^{\text{no-CSI}}$ , can be defined in analogous fashion with the only difference being in the definitions of the encoding and decoding functions. In particular, when there is perfect CSI, we have

$$X_i(t) = f_{i,t}^{(n)}\left(\mathcal{M}_i, \overline{H}(t)\right) \text{ and } \hat{\mathcal{M}}_i = g_i^{(n)}\left(\overline{Y}_i(n), \overline{H}(n)\right).$$

When there is no CSI, we have

$$X_i(t) = f_{i,t}^{(n)}\left(\mathcal{M}_i\right) \text{ and } \hat{\mathcal{M}}_i = g_i^{(n)}\left(\overline{Y}_i(n), \{H_{i1}(t), H_{i2}(t)\}_{t=1}^n\right).$$

Clearly,  $\mathbf{D}^{\text{no-CSI}} \subseteq \mathbf{D}^{\text{d-CSI}} \subseteq \mathbf{D}^{\text{p-CSI}}$ .

### III. THE DOF REGION OF THE IC WITH DELAYED CSI

In this section the main results regarding the DoF region of the MIMO IC are stated for which we need the following two definitions.

**Definition 1:** For a given  $i \in \{1, 2\}$ , Condition  $i$  is said to hold, whenever the inequality

$$M_i > N_1 + N_2 - M_j > N_i > N_j > M_j > N_j \frac{N_j - M_j}{N_i - M_j}$$

is true for  $j \in \{1, 2\}$  with  $j \neq i$ .

Clearly, the two conditions are symmetric counterparts of each other (i.e., one can be obtained from the other by switching user indices 1 and 2). Moreover, the two conditions can not be true simultaneously, and Condition  $i$  can not hold if  $N_j \geq N_i$ .

**Definition 2:** The region  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is defined as follows:

$$\begin{aligned} \mathbf{D}_{\text{outer}}^{\text{d-CSI}} = & \left\{ (d_1, d_2) \mid L_{o1} \equiv 0 \leq d_1 \leq \min(M_1, N_1) \text{ and } L_{o2} \equiv 0 \leq d_2 \leq \min(M_2, N_2), \right. \\ & L_1 \equiv \frac{d_1}{\min(N_1 + N_2, M_1)} + \frac{d_2}{\min(N_2, M_1)} \leq \frac{\min(N_2, M_1 + M_2)}{\min(N_2, M_1)}; \\ & L_2 \equiv \frac{d_1}{\min(N_1, M_2)} + \frac{d_2}{\min(N_1 + N_2, M_2)} \leq \frac{\min(N_1, M_1 + M_2)}{\min(N_1, M_2)}; \\ & L_3 \equiv d_1 + d_2 \leq \min \{ M_1 + M_2, N_1 + N_2, \max(M_1, N_2), \max(M_2, N_1) \}; \\ & \text{if Condition 1 holds, } L_4 \equiv d_1 + d_2 \frac{N_1 + 2N_2 - M_2}{N_2} \leq N_1 + N_2; \\ & \left. \text{if Condition 2 holds, } L_5 \equiv d_2 + d_1 \frac{N_2 + 2N_1 - M_1}{N_1} \leq N_1 + N_2 \right\}. \end{aligned}$$

In the sequel, the first two “single-user” bounds on  $d_1$  and  $d_2$  appearing in the above definition are referred to as  $L_{o1}$  and  $L_{o2}$ , respectively; while the remaining five bounds on the weighted sum of  $d_1$  and  $d_2$  are referred to respectively as  $L_1, L_2, \dots, L_5$ .

The following theorem gives an outer-bound to the DoF region.

**Theorem 1 (Outer-bound):** The DoF region of the MIMO IC with delayed CSI is outer-bounded by  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ , i.e.,

$$\mathbf{D}^{\text{d-CSI}} \subseteq \mathbf{D}_{\text{outer}}^{\text{d-CSI}}.$$

*Proof:* If  $(d_1, d_2) \in \mathbf{D}^{\text{d-CSI}}$ , then  $d_i \leq \min(M_i, N_i)$  for each  $i$  because the DoF of the point-to-point MIMO channel with  $M$  transmit and  $N$  receive antennas are upper-bounded by  $\min(M, N)$  [27]. Now, note that if  $(d_1, d_2) \in \mathbf{D}^{\text{p-CSI}}$ , then  $d_1$  and  $d_2$  satisfy bound  $L_3$  [13] (see also [16, Section III-C]), and hence, if  $(d_1, d_2) \in \mathbf{D}^{\text{d-CSI}} \subseteq \mathbf{D}^{\text{p-CSI}}$ , then the bound  $L_3$  must hold. It now remains to prove the bounds  $L_1, L_2, L_4$  and  $L_5$ . To this end, we observe that bounds  $L_1$  and  $L_2$  are symmetric counterparts of each other, and so are the bounds  $L_4$  and  $L_5$ . Hence, it is sufficient to prove bounds  $L_1$  and  $L_4$ . The proofs of  $L_1$  and  $L_4$  as being the outer-bounds are provided in Sections IV and V, respectively. ■

The following theorem asserts that the above outer-bound is tight for all possible values of  $(M_1, M_2, N_1, N_2)$ .

**Theorem 2 (The DoF region):** The DoF region of the MIMO IC with delayed CSI is equal to  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . In other words,

$$\mathbf{D}^{\text{d-CSI}} = \mathbf{D}_{\text{outer}}^{\text{d-CSI}}.$$



*Proof:* It has been proved in the previous theorem that  $\mathbf{D}^{\text{d-CSI}} \subseteq \mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . It will be shown here that  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}} \subseteq \mathbf{D}^{\text{d-CSI}}$ , i.e., the region  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is achievable. In the remainder of the proof it is assumed, without loss of generality, that  $N_1 \geq N_2$  (thus Condition 2 can not hold).

We now divide the analysis into three main cases which are further subdivided to obtain a total of seven cases, which are analyzed individually in their respective sections (stated beside the definition of the case). Note that the assumption of  $N_1 \geq N_2$  applies to every case.

- Case 0:  $N_2 \geq M_1$  – See Section VI.
- Case A:  $M_1 > N_2$  and  $M_2 \geq N_2$ :
  - Case A.I:  $M_2 \geq N_1$  – See Section VI. <sup>2</sup>
  - Case A.II:  $M_2 < N_1$  – See Section VII.
- Case B:  $M_1 > N_2$  and  $N_2 > M_2$ :
  - Case B.0:  $N_1 = N_2$  – See Section VIII.
  - Case B.I:  $N_1 > N_2$  and  $N_1 \geq M_1$  – See Section VIII.
  - Case B.II:  $M_1 > N_1 > N_2$  and Condition 1 does not hold – See Section IX.
  - Case B.III:  $M_1 > N_1 > N_2$  and Condition 1 holds – See Section X.

It turns out that for some values of the 4-tuple  $(M_1, M_2, N_1, N_2)$ , we have  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ , which implies the achievability of the outer-bound, and hence the theorem. However, if  $\mathbf{D}^{\text{no-CSI}} \neq \mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ , we need to develop a new achievability scheme, which is done here using the idea of interference alignment, to prove that  $\mathbf{D}^{\text{d-CSI}} = \mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . The reader is referred to Sections VI - X for detailed proofs. ■

From the above two theorems, we obtain two corollaries corresponding to two different sets of assumptions about the availability of CSI. Consider first the case wherein it is assumed that both the receivers have perfect and instantaneous knowledge of all channel matrices while the transmitters (as before) know the channel matrices perfectly, but with a delay of one time unit. We refer to this case as “delayed CSI at the transmitters” and denote the corresponding DoF region by  $\mathbf{D}^{\text{d-CSI-T}}$ . The following corollary characterizes this DoF region.

**Corollary 1:** The DoF region of the MIMO IC with delayed CSI at the transmitters is equal

<sup>2</sup>Thus, under Case A.I, the inequality  $M_2 \geq N_1$  holds, in addition to the inequality  $N_1 \geq N_2$  and the inequalities that are true with Case A.

to  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ , i.e.,

$$\mathbf{D}^{\text{d-CSI-T}} = \mathbf{D}^{\text{d-CSI}}.$$

*Proof:* Since  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}} = \mathbf{D}^{\text{d-CSI}} \subseteq \mathbf{D}^{\text{d-CSI-T}}$ , the region  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is an inner-bound to  $\mathbf{D}^{\text{d-CSI-T}}$ . Therefore, it is sufficient to prove that  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is also an outer-bound. To this end, note that the bounds  $L_{o1}$ ,  $L_{o2}$ ,  $L_3$  are outer-bounds, even in the present case. Moreover, from Sections IV and V, we observe that the bounds  $L_1$  and  $L_4$  have been derived by assuming that both the receivers have perfect and instantaneous CSI of all channel matrices, and hence, they are applicable to the IC with delayed CSI, as well. ■

Consider next a different assumption about CSI knowledge. The transmitters, instead of having delayed CSI of all channel matrices, have delayed CSI of only the cross channel matrices. That is, at time  $t$ , the  $i^{\text{th}}$  transmitter knows perfectly the channel matrix  $H_{ji}(t-1)$  with  $j \neq i$ , but does not have any knowledge of the realizations of the other three channel matrices. The receivers, on the other hand, have delayed CSI (as in the last section), in addition to the instantaneous CSI of their incoming channel matrices. We refer to this case as “delayed CSI of cross channel matrices” and denote the corresponding DoF region by  $\mathbf{D}^{\text{d-CSI-c}}$ , which is characterized by the following corollary.

**Corollary 2:** The DoF region of the MIMO IC with delayed CSI of cross channel matrices is equal to  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ , i.e.,

$$\mathbf{D}^{\text{d-CSI-c}} = \mathbf{D}^{\text{d-CSI}}.$$

*Proof:* Since  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}} = \mathbf{D}^{\text{d-CSI}} \supseteq \mathbf{D}^{\text{d-CSI-c}}$ , the region  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is an outer-bound to  $\mathbf{D}^{\text{d-CSI-c}}$ . It is thus sufficient to show that it is achievable even if the transmitters have just delayed CSI of cross channel matrices. To this end, note that all achievability schemes developed in Sections VI-X (which prove the achievability of  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  with delayed CSI) make use of only delayed CSI of cross channel matrices at the transmitters. Hence, the schemes developed in these sections are directly applicable to the case of delayed CSI of cross channel matrices, which proves the achievability part. ■

### A. Summary of Results

We know from the previous section that  $\mathbf{D}^{\text{d-CSI}} = \mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . The outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is defined in terms of two single-user bounds and five bounds labeled as  $L_1$ ,  $L_2$ ,  $\dots$ ,  $L_5$ . In this section,

Case	Definition of the Case ( $N_1 \geq N_2$ )	Bounds Active	Reference	Remarks
0	$N_2 \geq M_1$	$L_3$	Lemma 8	$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$
Case A: $M_1 > N_2$ and $M_2 \geq N_2$				
A.I.	$M_2 \geq N_1$	$L_{\{1,2\}}$	Lemma 9	
1	$M_1 \leq N_1$	$L_1$	See	$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
2	$M_1 > N_1$ and $N_2 = M_2$	$L_2$ or $L_3$	Section	$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$
3	$M_1 > N_1$ and $N_2 < M_2$	$L_{\{1,2\}}$	VI.	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
A.II.	$M_2 < N_1$	$L_{\{1,3\}}$	Lemma 10	
1	$N_1 \geq M_1$	$L_1$	See Section	$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
2	$N_1 < M_1$	$L_{\{1,3\}}$	VII.	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
Case B: $M_1 > N_2$ and $N_2 > M_2$				
B.0	$N_1 = N_2$	$L_3$	Lemma 11	$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$
B.I.	$N_1 > N_2$ and $N_1 \geq M_1$	$L_1$	Lemma 12	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
B.II.	$M_1 > N_1 > N_2$ and Condition 1 does not hold	$L_{\{1,3\}}$	Lemma 13	
1	$M_2 \leq m$	$L_3$	See Section	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$
2	$M_1 = M'_1$ and $M_2 > m$	$L_{\{1,3\}}$	IX.	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
B.III.	$M_1 > N_1 > N_2$ and Condition 1 holds	$L_{\{1,3,4\}}$	Lemma 17	
1	$M_1 \geq N_1 + N_2 - m$	$L_{\{3,4\}}$	See Section	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$
2	$M_1 < N_1 + N_2 - m$	$L_{\{1,3,4\}}$	X.	$\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$

$$M'_1 = \min(M_1, N_1 + N_2 - M_2) \quad m = N_2 \frac{M'_1 - N_1}{M'_1 - N_2}$$

Condition 1 can equivalently be stated as  $M_1 > M'_1 > N_1 > N_2 > M_2 > m$ ; See Lemma 20 in Appendix A.

TABLE I

SUMMARY OF RESULTS ON THE DoF REGION OF THE IC WITH DELAYED CSI:  $N_1 \geq N_2$ .

we discuss how the shape of the DoF region changes with the number of antennas at the four terminals. To this end, we need to determine which of the bounds in the description of the outer-bound are essential (i.e., not redundant or not implied by the other bounds). Thus, it is convenient to have the following definition.

Let  $S$  be a subset of  $\{1, 2, 3, 4, 5\}$  and  $S^c$  be the complement of  $S$ , i.e.,  $S^c = \{1, 2, 3, 4, 5\} \setminus S$ ,

**Definition 3:** The bounds  $L_S$  are said to be active if the single-user bounds  $L_{o1}$  and  $L_{o2}$ , and  $\{L_i\}_{i \in S}$  imply bound(s)  $\{L_j\} \forall j \in S^c$  (bound  $L_{3+i}$  is trivially implied by the other when Condition 1 does not hold).

Situation	Cases
$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$	0, A.I.2, B.0
$\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} \neq \mathbf{D}^{\text{p-CSI}}$	A.I.1, A.II.1
$\mathbf{D}^{\text{no-CSI}} \neq \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$	B.II.1
$\mathbf{D}^{\text{no-CSI}} \neq \mathbf{D}^{\text{d-CSI}} \neq \mathbf{D}^{\text{p-CSI}}$	A.I.3, A.II.2, B.I, B.II.2, B.III.1, B.III.2

TABLE II

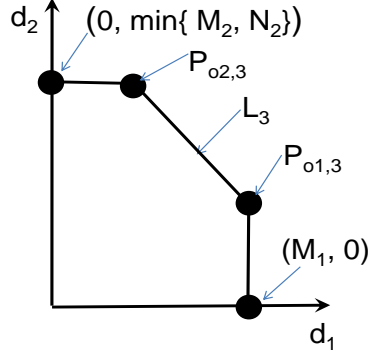
COMPARISON OF DoF REGIONS OF THE IC WITH NO, DELAYED, AND PERFECT CSI:  $N_1 \geq N_2$ .

Fig. 1. Typical Shape of the DoF Region of the 2-user MIMO IC with Delayed CSI: Case 0

If bound(s)  $L_S$  are active for some set  $S \subseteq \{1, 2, 3\}$ , then bound(s)  $\{L_j\}_{j \in S^c}$  can be dropped from the description of  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . Moreover, if bounds  $L_{\{S\}}$  are active with  $S = i$ , then we simply say that  $L_i$  is active.

Now, the shape of the DoF region can be determined by identifying the bounds that are active. Again, without loss of generality, it is assumed that  $N_1 \geq N_2$  (the results for  $N_2 > N_1$  can be obtained by switching the order of the two users).

The summary of our results on the DoF region of the MIMO IC with delayed CSI is presented in Table I. Various details can be easily obtained from this table. For example, under Case A.I, the inequality  $M_2 \geq N_1$  holds, in addition to the inequalities that define Case A. Moreover, as proved by Lemma 9, in this case, bounds  $L_{\{1,2\}}$  are active in general. But, more specifically, under Case A.I.1, only bound  $L_1$  is active (see the proof of Lemma 9 in Section VI) and  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} \neq \mathbf{D}^{\text{p-CSI}}$ ; whereas under Case A.I.2, we have  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$ , i.e.,  $L_3$  is active. In this manner information about all other cases can be obtained from Table I.

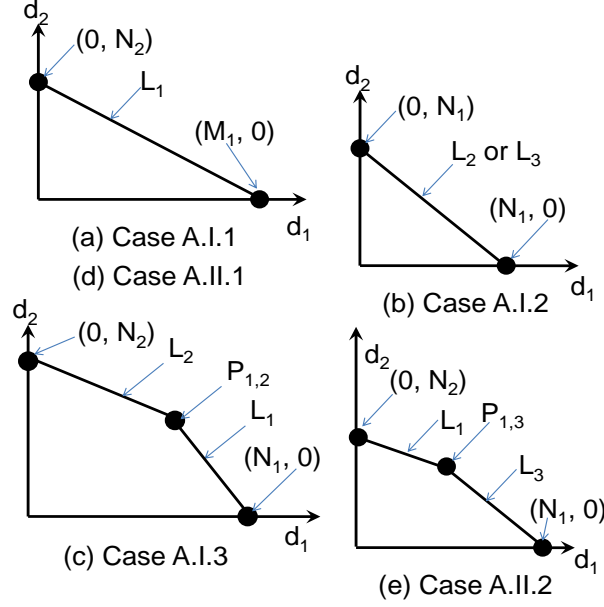


Fig. 2. Typical Shape of the DoF Region of the 2-user MIMO IC with Delayed CSI: Case A

In Figs. 1-3, we depict the DoF region of the IC with delayed CSI for the cases considered in Table I. Here, we make use of two facts: First, under Cases A and B,  $d_1 = \min(M_1, N_1) \Rightarrow d_2 = 0$ ; see Lemma 18. Second, under Case A,  $d_2 = \min(M_2, N_2) \Rightarrow d_1 = 0$ , which is not true with Case B; see Lemma 19.

### B. Comparison of the DoF Regions with No, Delayed, and Perfect CSI

For easy reference, we state the DoF regions of the IC with perfect and no CSI in the following remark.

**Remark 1:** The perfect-CSI DoF region of the IC is equal to the region  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  with bounds  $L_1$ ,  $L_2$ ,  $L_4$ , and  $L_5$  dropped [12] (see also [16, Subsection III-C]). Further, the no-CSI DoF region of the IC with  $N_1 \geq N_2$  is given by [14]–[16], [28]

$$\mathbf{D}^{\text{no-CSI}} = \left\{ (d_1, d_2) \mid 0 \leq d_i \leq \min(M_i, N_i) \text{ for } i = 1, 2 \right. \\ \left. \frac{d_1}{\min(M_1, N_1)} + \frac{d_2}{\min(M_1, N_2)} \leq \frac{\min(N_2, M_1 + M_2)}{\min(M_1, N_2)} \right\}.$$

Table I makes it easy to compare the DoF regions under the three cases of no, delayed, and perfect CSI. For example, we see that  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}}$  under Cases 0, A.I.1, A.I.2, A.II.1,

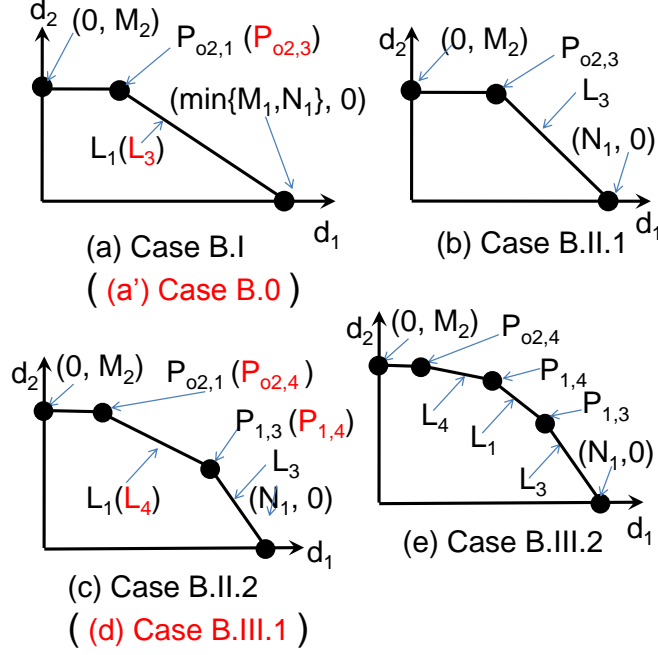


Fig. 3. Typical Shape of the DoF Region of the 2-user MIMO IC with Delayed CSI: Case B

and B.0. Moreover, out of these, under Cases 0, A.I.2, and B.0,  $\mathbf{D}^{\text{no-CSI}}$  equals  $\mathbf{D}^{\text{d-CSI}}$  because the perfect-CSI DoF region can be achieved even without CSI. In other words, under Cases A.I.1 and A.II.1, the availability of delayed CSI can not improve the no-CSI DoF region, even though the no CSI and perfect CSI DoF regions are not equal. These two cases can collectively be described by the inequalities  $N_1 \geq M_1 > N_2$  and  $M_2 \geq N_2$ .

Consider now the comparison of the delayed-CSI DoF region with the perfect-CSI DoF region. The two are equal under Cases 0, A.I.2, B.0, and B.II.1. Again, out of these, only under Case B.II.1, we have  $\mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}} \neq \mathbf{D}^{\text{no-CSI}}$ . These observations are summarized in Table II.

#### IV. PROOF OF THEOREM 1: $L_1$ IS AN OUTER-BOUND

Before starting the proof, we introduce the notation used in the remainder of the paper.

*Notation:* The set of four channel matrices at time  $t$  is denoted by  $H(t)$ , i.e.,  $H(t) = \{H_{ij}(t)\}$  where  $i, j \in \{1, 2\}$ . For integers  $n_1$  and  $n_2$ , if  $n_1 \leq n_2$ ,  $[n_1 : n_2] = \{n_1, n_1 + 1, \dots, n_2\}$ ; whereas if  $n_1 > n_2$ , then  $[n_1 : n_2]$  denotes the empty set. For a random variable  $X(t)$ ,  $X([n_1 : n_2]) = \{X(t)\}_{t=n_1}^{n_2}$  if  $n_1 \leq n_2$ , whereas  $X([n_1 : n_2])$  denotes an empty set if  $n_1 > n_2$ . Further, for  $n \geq 1$ ,  $\bar{X}(n) = X([1 : n])$ . For the received signal  $Y_i(t)$  and the channel matrix  $H_{ik}(t)$ , the  $j^{\text{th}}$

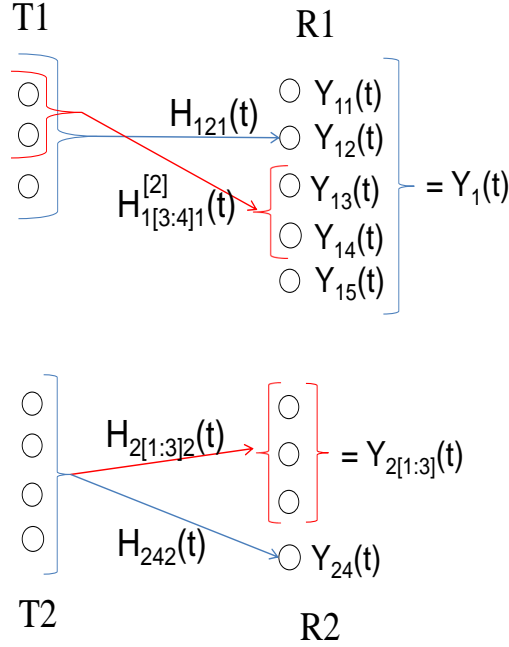


Fig. 4. Illustrating the Notation Used in This Paper

entry and the  $j^{th}$  row are denoted respectively by  $Y_{ij}(t)$  and  $H_{ijk}(t)$ . Further, whenever  $n_1 \leq n_2$  and  $n_3 \leq n_4$ ,  $Y_{i[n_1:n_2]}(t) = \{Y_{ij}(t)\}_{j=n_1}^{n_2}$ ,  $Y_{i[n_1:n_2]}([n_3:n_4]) = \{\{Y_{ij}(t)\}_{j=n_1}^{n_2}\}_{t=n_3}^{n_4}$ ,  $H_{i[n_1:n_2]j}(t)$  is the channel matrix from  $j^{th}$  transmitter to channel outputs  $Y_{i[n_1:n_2]}(t)$  (see Fig. 4); however, if  $n_1 > n_2$  and/or  $n_3 > n_4$ , then  $Y_{i[n_1:n_2]}(t)$  and  $Y_{i[n_1:n_2]}([n_3:n_4])$  denote empty sets. Moreover, for  $n \geq 1$ ,  $\bar{Y}_{i[n_1:n_2]}(n) = Y_{i[n_1:n_2]}([1:n])$ . Finally,  $o(\log_2 P)$  denotes any real-valued function  $x(P)$  of  $P$  such that  $\lim_{P \rightarrow \infty} \frac{x(P)}{\log_2 P} = 0$ .

As discussed earlier, the aim is to prove that  $L_1$  is a valid outer-bound. It turns out that to prove this bound, we need an inequality which is stated in the following lemma.

**Lemma 1:** Let  $m_1 = \min(M_1, N_1 + N_2)$  and  $m_2 = \min(M_1, N_2)$ . Then, for each  $n$ , we have

$$\frac{1}{m_2} h(\bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) \geq \frac{1}{m_1} h(\bar{Y}_1(n), \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) + n \cdot o(\log_2 P), \text{ or equivalently,}$$

$$\frac{1}{m_2} I(\mathcal{M}_1; \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) \geq \frac{1}{m_1} I(\mathcal{M}_1; \bar{Y}_1(n), \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) + n \cdot o(\log_2 P),$$

where the term  $o(\log_2 P)$  is constant with  $n$ .

In fact, the proof of bound  $L_1$  is an application of the above lemma. In the following section, bound  $L_1$  is obtained using the above lemma. Later, in Section IV-B, Lemma 1 is proved. Section

IV-C contains some remarks on the proof.

#### A. Proof of Bound $L_1$

We first outer-bound the capacity region of the given MIMO IC with delayed CSI as follows. Assume that both the receivers know all channel matrices perfectly and instantaneously. Further, R1 is assumed have non-causal knowledge of the message  $\mathcal{M}_2$  and instantaneous knowledge of the signal  $Y_2(t)$ . These assumptions can only enhance the capacity region of the given IC. We apply Fano's inequality [29] under these assumptions to upper bound the rates achievable for the two users as follows:

$$\begin{aligned} nR_2 &\leq I(\mathcal{M}_2; \bar{Y}_2(n) | \bar{H}(n)) + n\epsilon_n \\ &= h(\bar{Y}_2(n) | \bar{H}(n)) - h(\bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) + n\epsilon_n \end{aligned} \quad (3)$$

$$= \underbrace{I(\mathcal{M}_1, \mathcal{M}_2; \bar{Y}_2(n) | \bar{H}(n))}_{=t_1} - \underbrace{I(\mathcal{M}_1; \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n))}_{=t_2} + n\epsilon_n \text{ and} \quad (4)$$

$$nR_1 \leq \underbrace{I(\mathcal{M}_1; \bar{Y}_1(n), \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n))}_{=t_3} + n\epsilon_n \quad (5)$$

where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Let  $f(\cdot)$  denote the operation:

$$f(x) = \lim_{P \rightarrow \infty} \left\{ \frac{1}{\log_2 P} \cdot \lim_{n \rightarrow \infty} \frac{x}{n} \right\}$$

whenever the limits exist. Note here that the operator  $f(\cdot)$  preserves inequalities, i.e., if  $x \leq y$ , then  $f(x) \leq f(y)$ .

We now apply  $f(\cdot)$  to both the sides of equations (4) and (5) to obtain

$$d_2 \leq f(t_1) - f(t_2) \text{ and } d_1 \leq f(t_3), \quad (6)$$

where we make use of the facts that  $f(nR_i) = d_i$  for  $i = 1, 2$ ,  $f(n\epsilon_n) = 0$ .

Consider the first term  $f(t_1)$ , which stands for the DoF available at R2 per time slot (or the total number of dimensions of the receive-signal space of R2 per time slot). Since the DoF of the point-to-point MIMO channel are limited by the minimum of the number of transmit and receive antennas, we get

$$f(t_1) \leq \min(M_1 + M_2, N_2). \quad (7)$$



Consider the term  $f(t_2)$ . Given the message  $\mathcal{M}_2$  and the channel matrices  $\overline{H}(n)$ , the transmit signal  $\overline{X}_2(n)$  is deterministic. Hence, the term  $t_2$  equals the amount of information R2 can get about  $\mathcal{M}_1$  after decoding its intended message  $\mathcal{M}_2$ . That is, the term  $t_2$  is a measure of the interference caused by the transmission of T1 at R2, and  $f(t_2)$  is equal to the DoF occupied by the interference at R2 per time slot. Thus, the first inequality in (6) says that the number of DoF that can be achieved for the second user equals the total number of DoF available at R2 per time slot minus the number of DoF occupied by the interference. Moreover, using Lemma 1 and the second inequality in (6), we get

$$f(t_2) \geq \frac{m_2}{m_1} f(t_3) \geq \frac{m_2}{m_1} d_1, \quad (8)$$

which shows that if  $d_1$  DoF are to be achieved for the first user when there is delayed CSI, interference will occupy at least  $\frac{m_2}{m_1}$  times  $d_1$  DoF at R2, regardless of the achievability scheme used. This lower-bound on the DoF of the interference allows us to obtain an upper-bound on  $d_2$  as follows:

$$d_2 \leq \min(N_2, M_1 + M_2) - \frac{m_2}{m_1} d_1, \quad (9)$$

where we make use of inequalities (6), (7), and (8), and the above bound in turn yields us

$$\frac{d_1}{m_1} + \frac{d_2}{m_2} \leq \frac{\min(N_2, M_1 + M_2)}{m_2},$$

which is the bound  $L_1$ . ■

### B. Proof of Lemma 1

We will derive a series of lemmas and a corollary using which Lemma 1 will be proved. Let us first focus on the first inequality, using which the second inequality will be proved later. The first inequality relates the two differential entropy terms involving random variables  $\overline{Y}_1(n)$  and  $\overline{Y}_2(n)$ , which denote the signal received by the two receivers over the blocklength of  $n$ . In the following two lemmas, it is shown that although the received signals  $Y_1(t)$  and  $Y_2(t)$  are  $N_1$  and  $N_2$  dimensional, respectively, only the first  $m_1 - m_2$  and  $m_2$  entries of them are relevant as far as the current DoF analysis is concerned.

**Lemma 2:** If  $m_2 = \min(M_1, N_2)$ , then we have following:

$$h(\overline{Y}_2(n) | \mathcal{M}_2, \overline{H}(n)) \geq h(\overline{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \overline{H}(n)) + n \cdot o(\log_2 P),$$

where the term  $o(\log_2 P)$  is constant with  $n$ .

*Proof:* This can be proved as follows:

$$\begin{aligned}
& h(\bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) \\
&= h(\bar{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) + h(\bar{Y}_{2[m_2+1:N_2]}(n) | \bar{Y}_{2[1:m_2]}(n), \mathcal{M}_2, \bar{H}(n)) \\
&\geq h(\bar{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) + h(\bar{Y}_{2[m_2+1:N_2]}(n) | \bar{X}_1(n), \bar{X}_2(n), \bar{Y}_{2[1:m_2]}(n), \mathcal{M}_2, \bar{H}(n)) \quad (10) \\
&= h(\bar{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) + h(\bar{W}_{2[m_2+1:N_2]}(n)) \quad (11) \\
&= h(\bar{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) + n \cdot o(\log_2 P), \quad (12)
\end{aligned}$$

where inequality (10) follows because conditioning reduces entropy; the next equality (11) holds because translation does not change differential entropy and because noise is independent of the channel matrices, the transmit signals, and the messages; and the last equality is true since noise random variables are i.i.d. across time and their statistics are constant with  $P$ , which also explains why the term  $o(\log_2 P)$  is constant with  $n$ . ■

The following lemma deals with the second differential term involved in the first inequality of Lemma 1.

**Lemma 3:** If  $m_1 = \min(M_1, N_1 + N_2)$ , then

$$h(\bar{Y}_1(n), \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n)) \leq h(\bar{Y}_{1[1:m_1-m_2]}(n), \bar{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) + n \cdot o(\log_2 P).$$

where the term  $o(\log_2 P)$  is constant with  $n$ .

*Proof:* This lemma holds trivially if  $M_1 \geq N_1 + N_2$ . We thus consider the case where  $M_1 < N_1 + N_2$ . Given  $\mathcal{M}_2$  and  $\bar{H}(n)$ , the transmit signal  $X_2(t)$  is deterministic  $\forall t \in [1 : n]$ . Define  $Y'_i(t) = Y_i(t) - H_{i2}(t)X_2(t)$  and then define quantities  $\bar{Y}'_i(n)$ ,  $\bar{Y}'_{2[1:m_2]}(n)$ , etc. in an analogous fashion. Since translation does not change differential entropy, we get

$$h(\bar{Y}_{1[1:m_1-m_2]}(n), \bar{Y}_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) = h(\bar{Y}'_{1[1:m_1-m_2]}(n), \bar{Y}'_{2[1:m_2]}(n) | \mathcal{M}_2, \bar{H}(n)) =: q.$$

Denote by  $Q(n)$  random variables  $\{\mathcal{M}_2, \bar{H}(n)\}$ . Then we have

$$\begin{aligned}
& h(\bar{Y}_1(n), \bar{Y}_2(n) | Q(n)) = h(\bar{Y}'_1(n), \bar{Y}'_2(n) | Q(n)) \\
&= q + h(\bar{Y}'_{1[m_1-m_2+1:N_1]}(n), \bar{Y}'_{2[m_2+1:N_2]}(n) | Q(n), \bar{Y}'_{1[1:m_1-m_2]}(n), \bar{Y}'_{2[1:m_2]}(n)) \\
&\leq q + \sum_{t=1}^n q_t \text{ with } q_t = h(Y'_{1[m_1-m_2+1:N_1]}(t), Y'_{2[1:m_2]}(t) | H(t), Y'_{1[1:m_1-m_2]}(t), Y'_{2[1:m_2]}(t)),
\end{aligned}$$

where the inequality follows because the conditioning reduces entropy. Now, given the signal received at  $m_1$  receive antennas, namely,  $Y'_{1[1:m_1-m_2]}(t)$  and  $Y'_{2[1:m_2]}(t)$ , a noisy version of  $X_1(t)$  can be constructed via channel inversion with probability 1 (c.f. [16, Section II-C]). Hence, the term  $q_t$  just equals the differential entropy of some noise random variables. Since noise as well as channel matrices are i.i.d. across time and the statistics of both are constant with respect to  $P$ , the terms  $q_t$  are equal for all  $t$  and  $q_t = o(\log_2 P)$ , irrespective of  $n$ . ■

**Remark 2:** Since all the differential entropy or the mutual information terms considered in this section are conditioned on  $\mathcal{M}_2$  and  $\bar{H}(n)$ , it is assumed henceforth in this section that  $X_2(t) = 0, \forall t$  because  $H_{ij2}(t)X_2(t)$  can always be subtracted from the received signal  $Y_{ij}(t)$ .

Consider the following lemma, which is the crucial step in establishing Lemma 1.

**Lemma 4:** For a given  $t \in [1 : n]$ ,  $i \in [1 : m_2 - 1]$ , and  $j = i + 1$ , we have

$$h\left(Y_{2i}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:i-1]}(t)\right) \geq h\left(Y_{2j}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:i]}(t)\right).$$

*Proof:* Consider the first term

$$\begin{aligned} & h\left(Y_{2i}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:i-1]}(t)\right) \\ &= h\left(Y_{2i}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), Y_{2[1:i-1]}(t), \mathcal{M}_2, \bar{H}(t-1), H_{2[1:i-1]1}(t)\right) \end{aligned} \quad (13)$$

$$= \mathbb{E}_{H_{2i1}(t)=a} h\left(Y_{2i}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), Y_{2[1:i-1]}(t), \mathcal{M}_2, \bar{H}(t-1), H_{2[1:i-1]1}(t), H_{2i1}(t) = a\right) \quad (14)$$

$$= \mathbb{E}_{H_{2j1}(t)=a} h\left(Y_{2j}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), Y_{2[1:i-1]}(t), \mathcal{M}_2, \bar{H}(t-1), H_{2[1:i-1]1}(t), H_{2j1}(t) = a\right) \quad (15)$$

$$\geq h\left(Y_{2j}(t) \middle| \bar{Y}_{2[1:m_2]}(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:i]}(t)\right), \quad (16)$$

where equality (13) holds because all the concerned random variables are independent of  $H_{2[i+1:N_2]1}(t)$ ; the equation (14) follows from the definition of conditional differential entropy; the next equality (15) holds because conditioned of random variables  $\{\bar{Y}_{2[1:m_2]}(t-1), Y_{2[1:i-1]}(t), \mathcal{M}_2, \bar{H}(t-1), \text{ and } H_{2[1:i-1]1}(t)\}$ , the joint distribution of the pair of random variables  $\{H_{2i1}(t), X(t)\}$  is identical to that of the pair  $\{H_{2j1}(t), X(t)\}$ ; and the final inequality (16) holds since conditioning reduces entropy. ■

Consider the equality in (15), which proves that the the signals received  $Y_{2i}(t)$  and  $Y_{2j}(t)$  at the  $i^{th}$  and  $j^{th}$  antenna, respectively, of R2 have equal differential entropy, when conditioned on the channel matrices  $\bar{H}(t)$ , the message  $\mathcal{M}_2$ , the past channel outputs  $\bar{Y}_2(t-1)$ , and the present channel outputs  $Y_{2[1:i-1]}(t)$  at some other receive antennas. We refer to this property as

the statistical equivalence of the channel outputs, which essentially says that given the past and present channel outputs, the signals received at any two antennas of the system provide equal amount of information about  $\mathcal{M}_1$  (when T2 is silent). This property provides a basis for us to relate the differential entropies of the signals received by R1 and R2, and indeed it is the important point of this proof. See Section IV-C for further discussion.

The following corollary uses the above lemma to obtain inequalities in more useful form.

**Corollary 3:** For a given  $t \in [1 : n]$ , given  $i \in [1 : m_1 - m_2 - 1]$ ,  $j = i + 1$ ,  $V_1(t) = \{\bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\}$ , and  $V_2(t) = \{\bar{Y}_{2[1:m_2]}(t), \bar{Y}_{1[1:m_1-m_2]}(t-1), \mathcal{M}_2, \bar{H}(t)\}$ , we have the following inequalities:

$$\begin{aligned} h\left(Y_{2[1:m_2]}(t) \middle| V_1(t)\right) &\geq m_2 \cdot h\left(Y_{2m_2}(t) \middle| V_1(t), Y_{2[1:m_2-1]}(t)\right) \\ h\left(Y_{2m_2}(t) \middle| V_1(t), Y_{2[1:m_2-1]}(t)\right) &\geq h\left(Y_{11}(t) \middle| V_2(t)\right), \\ h\left(Y_{1i}(t) \middle| V_2(t), Y_{1[1:i-1]}(t)\right) &\geq h\left(Y_{1j}(t) \middle| V_2(t), Y_{1[1:i]}(t)\right), \text{ and} \\ (m_1 - m_2) \cdot h\left(Y_{11}(t) \middle| V_2(t)\right) &\geq h\left(Y_{1[1:m_1-m_2]}(t) \middle| V_2(t)\right). \end{aligned}$$

*Proof:* By applying the previous lemma, we obtain

$$\begin{aligned} h\left(Y_{2m_2}(t) \middle| V_1(t), Y_{2[1:m_2-1]}(t)\right) &\leq h\left(Y_{2(m_2-1)}(t) \middle| V_1(t), Y_{2[1:m_2-2]}(t)\right) \\ &\leq h\left(Y_{2(m_2-2)}(t) \middle| V_1(t), Y_{2[1:m_2-3]}(t)\right) \leq \cdots \leq h\left(Y_{2k}(t) \middle| V_1(t), Y_{2[1:k-1]}(t)\right) \\ &\leq \cdots \leq h\left(Y_{22}(t) \middle| V_1(t), Y_{21}(t)\right) \leq h\left(Y_{21}(t) \middle| V_1(t)\right), \end{aligned}$$

which implies that

$$\begin{aligned} h\left(Y_{2[1:m_2]}(t) \middle| V_1(t)\right) &= \sum_{i=1}^{m_2} h\left(Y_{2i}(t) \middle| V_1(t), Y_{2[1:i-1]}(t)\right) \\ &\geq \sum_{i=1}^{m_2} h\left(Y_{2m_2}(t) \middle| V_1(t), Y_{2[1:m_2-1]}(t)\right) = m_2 \cdot h\left(Y_{2m_2}(t) \middle| V_1(t), Y_{2[1:m_2-1]}(t)\right). \end{aligned}$$

The next inequality is based on the idea of statistical equivalence of the signal received at two of the receive antennas. It can be proved as follows:

$$\begin{aligned} h\left(Y_{2m_2}(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:m_2-1]}(t)\right) &= h\left(Y_{11}(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:m_2-1]}(t)\right) \\ &\geq h\left(Y_{11}(t) \middle| V_2(t)\right), \end{aligned}$$

where the last inequality follows because conditioning reduces entropy. The third inequality is immediate from the previous lemma, whereas the last inequality can be proved in a manner analogous to the proof of the first inequality of this corollary. ■

The next lemma uses the above corollary.

**Lemma 5:** The following inequality holds:

$$\frac{1}{m_2} h\left(\bar{Y}_{2[1:m_2]}(n) \middle| \mathcal{M}_2, \bar{H}(n)\right) \geq \frac{1}{m_1} h\left(\bar{Y}_{1[1:m_1-m_2]}(n), \bar{Y}_{2[1:m_2]}(n) \middle| \mathcal{M}_2, \bar{H}(n)\right).$$

*Proof:* To simplify the notation, with abuse of notation, we omit in this proof the subscripts  $[1 : m_2]$  and  $[1 : m_1 - m_2]$  appearing in the above differential entropy terms. That is, we write  $Y_{2[1:m_2]}(t)$ ,  $Y_{1[1:m_1-m_2]}(t)$ ,  $\bar{Y}_{2[1:m_2]}(n)$ , and  $\bar{Y}_{1[1:m_1-m_2]}(n)$  respectively as  $Y_2(t)$ ,  $Y_1(t)$ ,  $\bar{Y}_2(n)$ , and  $\bar{Y}_1(n)$ . Then

$$\begin{aligned} h\left(\bar{Y}_2(n) \middle| \mathcal{M}_2, \bar{H}(n)\right) &= \sum_{t=1}^n h\left(Y_2(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right) \text{ and} \\ h\left(\bar{Y}_1(n), \bar{Y}_2(n) \middle| \mathcal{M}_2, \bar{H}(n)\right) &= \sum_{t=1}^n h\left(Y_1(t), Y_2(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right), \end{aligned}$$

where we make use of the fact that random variables  $\bar{Y}_1(t)$ ,  $\bar{Y}_2(t)$ ,  $\mathcal{M}_2$ , and  $\bar{H}(t)$  are independent of  $H([t+1 : n])$ . Hence it is sufficient to prove that for each  $t \in [1 : n]$ ,

$$m_1 \cdot h\left(Y_2(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right) \geq m_2 \cdot h\left(Y_1(t), Y_2(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right).$$

Moreover, since  $h\left(Y_1(t), Y_2(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right)$

$$= h\left(Y_2(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right) + h\left(Y_1(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t), \mathcal{M}_2, \bar{H}(t)\right)$$

and

$$h\left(Y_2(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right) \geq h\left(Y_2(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right),$$

it is enough to establish the following inequality:

$$(m_1 - m_2) \cdot h\left(Y_2(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right) \geq m_2 \cdot h\left(Y_1(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t), \mathcal{M}_2, \bar{H}(t)\right).$$

To this end, consider the following:

$$\begin{aligned} &(m_1 - m_2) \cdot h\left(Y_2(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t)\right) \\ &\geq (m_1 - m_2) \cdot m_2 \cdot h\left(Y_{2m_2}(t) \middle| \bar{Y}_2(t-1), \mathcal{M}_2, \bar{H}(t), Y_{2[1:m_2-1]}(t)\right) \\ &\geq m_2 \cdot (m_1 - m_2) \cdot h\left(Y_{11}(t) \middle| \bar{Y}_2(t), \bar{Y}_1(t-1), \mathcal{M}_2, \bar{H}(t)\right) \\ &\geq m_2 \cdot h\left(Y_1(t) \middle| \bar{Y}_1(t-1), \bar{Y}_2(t), \mathcal{M}_2, \bar{H}(t)\right), \end{aligned}$$

where the three inequalities follow respectively from the first, second, and the fourth inequalities of Corollary 3. ■

Using the above lemma, Lemma 1 can now be proved as follows.

*Proof of Lemma 1:* Consider the first inequality. Recall that the differential entropy term on the left hand side of the above inequality is lower-bounded (within  $o(\log_2 P)$ ) via Lemma 2, whereas the one on the right hand side is upper-bounded (within  $o(\log_2 P)$ ) via Lemma 3. The inequality now follows from Lemma 5 by noting that the sum or the difference of two  $o(\log_2 P)$  yields another  $o(\log_2 P)$  term.

The second inequality follows from the first by noting that the differential entropy of the received signals conditioned on the channel matrices and the messages is equal to that of some noise terms, which is of the order of  $n \cdot o(\log_2 P)$ . ■

### C. Comments on the Proof of Lemma 1

Let us revisit the equality in (15), which proves the statistical equivalence of the channel outputs at two of the receive antennas, namely,  $Y_{2i}(t)$  and  $Y_{2j}(t)$ , given  $\bar{H}(t)$ ,  $\mathcal{M}_2$ ,  $\bar{Y}(t-1)$  and  $Y_{2[1:i-1]}(t)$ . The random variables  $\{\bar{Y}_2(t-1), Y_{2[1:i-1]}(t), \mathcal{M}_2, \bar{H}(t-1), H_{2[1:i-1]1}(t)\}$  are dependent on the transmit signal  $X_1(t)$ . However, conditioned on this set of random variables, the transmit signal  $X_1(t)$  is independent of  $H_{2i1}(t)$  as well as  $H_{2j1}(t)$ , and moreover, the channel vectors are  $H_{2i1}(t)$  and  $H_{2j1}(t)$  are themselves i.i.d. Hence, conditioned on the above set of random variables, the situation at the  $i^{th}$  antenna of R2 is identical to that at its  $j^{th}$  antenna, and hence, they would provide equal amount of information about the message. This observation is the key to the equality in (15), which is important for proving Lemma 1.

Note here that the equality of the two differential entropy terms claimed in (15) holds because the random variables, on which these two terms are conditioned, involve only the past channel outputs and the present channel outputs at some other receive antennas of the system, in addition to the channel matrices and the message  $\mathcal{M}_2$ . In fact, the same result may not hold if the two differential entropy terms are conditioned on some future channel outputs as well. This is because these future channel outputs are correlated with the signal transmitted (by T1) during some future time slot, which, due to the availability of delayed CSI, can depend on the present channel outputs  $Y_{2i}(t)$  and  $Y_{2j}(t)$  (and hence on the channel vectors  $H_{2i1}(t)$  and  $H_{2j1}(t)$ ) in any arbitrary fashion. Hence, conditioned on the future channel outputs, the vectors  $H_{2i1}(t)$  and  $H_{2j1}(t)$  may no longer

be identically distributed, which otherwise is the case. This implies that the equality in (15) may not hold.

Recall that the outer-bound derived in [16] for the no-CSI case also rests on the statistical equivalence of the channel outputs. Particularly, in the case of no CSI, the channel outputs corresponding to any two receive antennas would have equal differential entropy, when conditioned on the same set of random variables, because the channel inputs are independent of all channel matrices. In other words, when there is no CSI, the statistical equivalence holds as long as the two differential entropies are conditioned on the same set of random variables, regardless of what these random variables are. In this sense, as compared to the case of delayed CSI, the property of statistical equivalence holds more generally under no CSI.

As mentioned earlier, an outer-bound to the DoF region of the  $K$ -user MISO BC with delayed CSI was derived in [20], which was later generalized by the authors in [22] to the case of MIMO BC. These outer-bounds have been found to be tight in some special cases (see [20], [22]) for details). These bounds have been derived by making use of the result of [24] which states that feedback can not improve the capacity region of the physically-degraded BC. However, using the idea developed here for proving bound  $L_1$ , it is possible to derive the outer-bound derived in [22] without making use of the result of [24].

## V. PROOF OF THEOREM 1: $L_4$ IS AN OUTER-BOUND

We first outer-bound the capacity region of the given MIMO IC with delayed CSI by assuming that each receiver knows all channel matrices perfectly and instantaneously. Under this assumption, using Fano's inequality, we upper-bound the rates achievable for the two users as follows:

$$nR_2 = \underbrace{I(\mathcal{M}_1, \mathcal{M}_2; \bar{Y}_2(n) | \bar{H}(n))}_{=t_1} - \underbrace{I(\mathcal{M}_1; \bar{Y}_2(n) | \mathcal{M}_2, \bar{H}(n))}_{=t_2} + n\epsilon_n \text{ and} \quad (17)$$

$$nR_1 \leq \underbrace{I(\mathcal{M}_1; \bar{Y}_1(n) | \bar{H}(n))}_{=t_3} + n\epsilon_n, \quad (18)$$

where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Using these bounds, we get

$$d_2 \leq N_2 - f(t_2) \quad (19)$$

$$d_1 \leq f(t_3) = f\left\{ I(\mathcal{M}_1; \bar{Y}_1(n) | \bar{H}(n)) \right\}, \quad (20)$$

where we use equation (7). In the following, we obtain a tight upper-bound on  $f(t_3)$  so that the desired bound  $L_4$  can be derived.

To this end, define a unitary matrix  $U_{12}(t)$  such that the last  $N_1 - M_2$  rows of  $U_{12}(t)H_{12}(t)$  consist only of zeros. Such a unitary matrix can be obtained from the singular-value decomposition [30] of  $H_{12}(t)$ , and therefore,  $U_{12}(t)$  is a deterministic function of  $H_{12}(t)$ . Define  $Y'_1(t) = U_{12}(t)Y_1(t)$ . Note that  $X_2(t)$  affects only the first  $M_2$  entries of  $Y'(t)$ . Since a unitary transformation can not affect the mutual information, we have

$$\begin{aligned} d_1 &\leq f\left\{I\left(\mathcal{M}_1; \bar{Y}'_1(n) \middle| \bar{H}(n)\right)\right\} \\ &= f\left\{I\left(\mathcal{M}_1; \bar{Y}'_{1[1:M_2]}(n) \middle| \bar{Y}'_{1[M_2+1:N_1]}(n), \bar{H}(n)\right)\right\} + f\left\{I\left(\mathcal{M}_1; \bar{Y}'_{1[M_2+1:N_1]}(n) \middle| \bar{H}(n)\right)\right\}. \end{aligned} \quad (21)$$

We upper-bound each of the two terms appearing above through the following two lemmas.

**Lemma 6:** We have

$$f\left\{I\left(\mathcal{M}_1; \bar{Y}'_{1[1:M_2]}(n) \middle| \bar{Y}'_{1[M_2+1:N_1]}(n), \bar{H}(n)\right)\right\} \leq M_2 - d_2. \quad (22)$$

*Proof:* See Section V-A. ■

**Lemma 7:** We have

$$f\left\{I\left(\mathcal{M}_1; \bar{Y}'_{1[M_2+1:N_1]}(n) \middle| \bar{H}(n)\right)\right\} \leq f\left\{I\left(\mathcal{M}_1; \bar{Y}'_{1[M_2+1:N_1]}(n) \middle| \mathcal{M}_2, \bar{H}(n)\right)\right\} \quad (23)$$

$$= f\left\{I\left(\mathcal{M}_1; \bar{Y}_{1[M_2+1:N_1]}(n) \middle| \mathcal{M}_2, \bar{H}(n)\right)\right\} \quad (24)$$

$$\leq f\left\{I\left(\mathcal{M}_1; \bar{Y}_{1[M_2+1:N_1]}(n), \bar{Y}_2 \middle| \mathcal{M}_2, \bar{H}(n)\right)\right\} \quad (25)$$

$$\leq \frac{N_1 + N_2 - M_2}{N_2} f\left\{I\left(\mathcal{M}_1; \bar{Y}_2 \middle| \mathcal{M}_2, \bar{H}(n)\right)\right\} = \frac{N_1 + N_2 - M_2}{N_2} f(t_2). \quad (26)$$

*Proof:* See Section V-B. ■

Using inequalities (21), (22), and (26), we obtain

$$d_1 \leq (M_2 - d_2) + \frac{N_1 + N_2 - M_2}{N_2} f(t_2) \Rightarrow f(t_2) \geq \frac{N_2}{N_1 + N_2 - M_2} (d_1 + d_2 - M_2).$$

We substitute this lower-bound on  $f(t_2)$  into equation (19), we get

$$\begin{aligned} \frac{d_2}{N_2} &\leq 1 - \frac{1}{N_1 + N_2 - M_2} (d_1 + d_2 - M_2) \\ &\Rightarrow \frac{d_1}{N_1 + N_2 - M_2} + d_2 \left\{ \frac{1}{N_2} + \frac{1}{N_1 + N_2 - M_2} \right\} \leq 1 + \frac{M_2}{N_1 + N_2 - M_2} \\ &\Rightarrow \frac{d_1}{N_1 + N_2 - M_2} + d_2 \frac{N_1 + 2N_2 - M_2}{N_2(N_1 + N_2 - M_2)} \leq \frac{N_1 + N_2}{N_1 + N_2 - M_2}, \end{aligned}$$

which is the desired bound  $L_4$ .



### A. Proof of Lemma 6

To prove this lemma, we make use of the techniques developed in [12]. We first define to following quantities: Let the matrix formed by retaining the first  $M_2$  rows of  $U_{12}(t)H_{12}(t)$  be  $Z(t)$  (it is thus  $M_2 \times M_2$  in size); define

$$\alpha(t) = \min \left\{ \frac{1}{\sigma_{\max}^2[Z(t)]}, \frac{1}{\sigma_{\max}^2[H_{22}(t)]} \right\},$$

where  $\sigma_{\max}[A]$  denotes the largest singular-value of  $A$ ; let  $W_b(t)$  be an  $M_2$ -dimensional noise vector that is distributed as

$$W_b(t) \sim \mathcal{CN} \left( 0, Z(t) \left\{ [Z(t)]^* Z(t) \right\}^{-1} [Z(t)]^* - \alpha(t) Z(t) [Z(t)]^* \right)$$

with its realizations being i.i.d. across time; and finally, set  $Y_{1[1:M_2]}^\dagger(t) = Y'_{1[1:M_2]}(t) - W_b(t)$ .

Now consider the following:

$$\begin{aligned} f \left\{ I \left( \mathcal{M}_1; \bar{Y}'_{1[1:M_2]}(n) \middle| \bar{Y}'_{1[M_2+1:N_1]}(n), \bar{H}(n) \right) \right\} &\leq f \left\{ I \left( \mathcal{M}_1; \bar{Y}_{1[1:M_2]}^\dagger(n) \middle| \bar{Y}'_{1[M_2+1:N_1]}(n), \bar{H}(n) \right) \right\} \\ &= f \left\{ I \left( \mathcal{M}_1, \mathcal{M}_2; \bar{Y}_{1[1:M_2]}^\dagger(n) \middle| \bar{Y}'_{1[M_2+1:N_1]}(n), \bar{H}(n) \right) \right\} \\ &\quad - f \left\{ I \left( \mathcal{M}_2; \bar{Y}_{1[1:M_2]}^\dagger(n) \middle| \mathcal{M}_1, \bar{Y}'_{1[M_2+1:N_1]}(n), \bar{H}(n) \right) \right\} \\ &\leq M_2 - f \left\{ I \left( \mathcal{M}_2; \bar{Y}_{1[1:M_2]}^\dagger(n) \middle| \mathcal{M}_1, \bar{H}(n) \right) \right\}, \end{aligned} \tag{27}$$

where the first inequality holds because, within the partial order of positive semi-definite matrices, the covariance matrix of  $W_b(t)$  is smaller than the identity matrix [12]; while the last inequality follows from the following reasons: (a) the number of DoF of the point-to-point MIMO channel are limited by the number of receive antennas [27], and (b) conditioned on  $\mathcal{M}_1$  and  $\bar{H}(n)$ ,  $X_1(n)$  is deterministic, which implies that  $\mathcal{M}_2$  and  $\bar{Y}_{1[1:M_2]}^\dagger(n)$  are independent of  $\bar{Y}'_{1[M_2+1:N_1]}(n)$ , conditioned on  $\mathcal{M}_1$  and  $\bar{H}(n)$ .

Now, using the arguments developed in [12] (see Steps 3-5 in the proof of Theorem 1 therein), it can be shown that

$$f \left\{ I \left( \mathcal{M}_2; \bar{Y}_{1[1:M_2]}^\dagger(n) \middle| \mathcal{M}_1, \bar{H}(n) \right) \right\} \geq f \left\{ I \left( \mathcal{M}_2; \bar{Y}_2(n) \middle| \mathcal{M}_1, \bar{H}(n) \right) \right\} \geq d_2.$$

Substituting the above lower-bound into (27), we obtain the desired inequality.

### B. Proof of Lemma 7

The first inequality in (23) follows by noting that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are independent.

To prove the next equality in (24), consider the following argument. Conditioned on  $\mathcal{M}_2$  and  $\overline{H}(n)$ ,  $\overline{X}_2(n)$  is deterministic. Further, since translation does not change differential entropy [29], conditioned on  $\mathcal{M}_2$  and  $\overline{H}(n)$ , we may let  $X_2(t) = 0 \ \forall t \leq n$ . Thus, in the following, it may be assumed, without loss of generality, that

$$Y_1(t) = H_{11}(t)X_1(t) + W_1(t) \text{ and } Y'_1(t) = U_{12}(t)H_{11}(t)X_1(t) + U_{12}(t)W_1(t). \quad (28)$$

Let  $H'_{11}(t) = U_{12}(t)H_{11}(t)$ . Since  $U_{12}(t)$  is independent of  $H_{11}(t)$  and  $H_{11}(t)$  is i.i.d. Rayleigh faded, we have  $H'_{11}(t) \sim H_{11}(t)$ . For similar reasons,  $U_{12}(t)W_1(t) \sim W_1(t)$ . Now, let  $H'(t)$  be a collection of channel matrices  $H'_{11}(t)$ ,  $H'_{12}(t)$ ,  $H'_{21}(t)$ , and  $H'_{22}(t)$ . Then  $H(t) \sim H'(t)$ .

Since  $U_{12}(t)$  is a deterministic function of  $H_{12}(t)$ , a one-to-one and onto mapping exists between  $H(t)$  and  $H'(t)$ , which gives us the following equality:

$$f\left\{I\left(\mathcal{M}_1; \overline{Y}'_{1[M_2+1:N_1]}(n) \middle| \mathcal{M}_2, \overline{H}(n)\right)\right\} = f\left\{I\left(\mathcal{M}_1; \overline{Y}'_{1[M_2+1:N_1]}(n) \middle| \mathcal{M}_2, \overline{H}'(n)\right)\right\}. \quad (29)$$

The quantity on the right hand side of the above equation is a function of the joint distribution of random variables

$$\left\{\mathcal{M}_1, \mathcal{M}_2, \overline{H}'(n), \overline{X}_1(n), \overline{Y}'_1(n)\right\},$$

which, in fact, is equal to that of the random variables

$$\left\{\mathcal{M}_1, \mathcal{M}_2, \overline{H}(n), \overline{X}_1(n), \overline{Y}_1(n)\right\}$$

because  $H'(t) \sim H(t)$ ,  $U_{12}(t)W_1(t) \sim W_1(t)$ , and equation (28) holds. This equality of the joint distributions implies that

$$f\left\{I\left(\mathcal{M}_1; \overline{Y}'_{1[M_2+1:N_1]}(n) \middle| \mathcal{M}_2, \overline{H}'(n)\right)\right\} = f\left\{I\left(\mathcal{M}_1; \overline{Y}_{1[M_2+1:N_1]}(n) \middle| \mathcal{M}_2, \overline{H}(n)\right)\right\}. \quad (30)$$

The desired equality in (24) can now be obtained by combining equations (29) and (30).

The inequality (25) holds trivially. The next inequality (26) is a simple application of Lemma 1, and the final equality follows from the definition of  $t_2$ .

## VI. PROOF OF THEOREM 2: CASES 0 AND A.I

Consider first Case 0 for which we have the following lemma.

**Lemma 8:** Under Case 0,  $\mathbf{D}^{\text{p-CSI}} = \mathbf{D}^{\text{d-CSI}}$ , which implies that bound  $L_3$  is active.

*Proof:* Under Case 0, the inequality  $N_1 \geq N_2 \geq M_1$  is true (recall  $N_1 \geq N_2$  throughout). Then by [16, Remark 17], we have  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{p-CSI}}$ , which yields  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}} = \mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . This also implies that  $L_3$  is active. ■

The remainder of this section will now deal with Case A.I, for which the defining inequalities can be stated as

$$M_2 \geq N_1 \geq N_2 \text{ and } M_1 > N_2,$$

where we account for  $N_1 \geq N_2$ , the inequalities  $M_1 > N_2$  and  $M_2 \geq N_2$  that hold with Case A, and the one that defines Case A.I. The following lemma proves that the proposed outer-bound is tight and also helps determine its shape.

**Lemma 9:** Under Case A.I, bounds  $L_{\{1,2\}}$  are active and the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is achievable. More specifically, under Case A.I,

1. if  $M_1 \leq N_1$ , then  $L_1$  is active and  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$ ;
2. if  $M_1 > N_1$  and  $N_2 = M_2$ , then  $L_2$  is active and  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$ ;
3. if  $M_1 > N_1$  and  $N_2 < M_2$  then  $L_{\{1,2\}}$  is active and  $\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$ .

*Proof:* Consider the region  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$  obtained from the region  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  by simply ignoring the bound  $L_3$ . Clearly,  $\mathbf{D}^{\text{d-CSI}} \subseteq \mathbf{D}_{\text{outer}}^{\text{d-CSI}} \subseteq \mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$ . Hence, to prove this lemma, it is sufficient to establish the achievability of  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$ , which is the goal of this proof.

Under Case A.I, bounds  $L_1$  and  $L_2$  are given by

$$L_1 \equiv \frac{d_1}{\min(N_1 + N_2, M_1)} + \frac{d_2}{N_2} \leq 1 \text{ and } L_2 \equiv \frac{d_1}{N_1} + \frac{d_2}{\min(N_1 + N_2, M_2)} \leq 1, \text{ respectively.}$$

We can now have three different scenarios, depending upon whether only one of  $L_1$  and  $L_2$  is active or both are active:

- Case A.I.1:  $M_1 \leq N_1 \Rightarrow M_2 \geq N_1 \geq M_1 > N_2$ :

Since  $\min(M_1, N_1 + N_2) = M_1 \leq N_1$  and  $N_2 \leq \min(M_2, N_1 + N_2)$ ,  $L_1$  implies  $L_2$  (i.e., if  $L_1$  is true,  $L_2$  is also true), or  $L_1$  is active. Moreover, it can be verified that  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}} = \mathbf{D}^{\text{no-CSI}}$  (see Remark 1), which implies the achievability of  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$  and the fact that  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}}$ . Hence, the lemma holds in this case.

- Case A.I.2:  $\{M_1 > N_1 \text{ and } N_2 = M_2\} \Rightarrow M_1 > N_1 = M_2 = N_2$ :

Since  $N_2 = M_2$  and  $M_2 \geq N_1 \geq N_2$ , we have the inequality  $M_1 > N_1 = M_2 = N_2$ . Since  $\min(M_1, N_1 + N_2) > N_1$  and  $N_2 = M_2$ ,  $L_2$  is active. It can be verified that in this case, the region  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$  with  $L_2$  active coincides with  $\mathbf{D}^{\text{no-CSI}}$  as well as  $\mathbf{D}^{\text{p-CSI}}$ . Hence, we get  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$  (see also [16, Remark 17]).

- Case A.I.3:  $\{M_1 > N_1 \text{ and } N_2 < M_2\}$ :

In this case, we have

$$M_1 > N_1, N_2; \text{ and } M_2 \geq N_1 \geq N_2 \text{ with either } M_2 \neq N_1 \text{ or } N_1 \neq N_2.$$

Here, both the bounds  $L_{\{1,2\}}$  are (strictly) active because  $\min(N_1 + N_2, M_1) > N_1$  but  $N_2 < \min(M_2, N_1 + N_2)$ , and hence no bound can imply the other. The typical shape of  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$  is as shown in Fig. 2(c), whence we deduce that if the point  $P_{1,2}$  (i.e., the point of intersection of  $L_1$  and  $L_2$ ) is known to be achievable, the entire outer-bound  $\mathbf{D}_{\text{outer},1}^{\text{d-CSI}}$  can be achieved via time sharing. Thus, to prove the lemma, it is sufficient to show the achievability of point  $P_{1,2}$ .

An achievability scheme for point  $P_{1,2}$ :

Point  $P_{1,2}$  is given by

$$P_{1,2} \equiv \left( \frac{N_1 \cdot M'_1 \cdot (M'_2 - N_2)}{N_1(M'_2 - N_2) + M'_2(M'_1 - N_1)}, \frac{N_2 \cdot M'_2 \cdot (M'_1 - N_1)}{N_1(M'_2 - N_2) + M'_2(M'_1 - N_1)} \right),$$

where<sup>3</sup>  $M'_i = \min(M_i, N_1 + N_2)$  for  $i = 1, 2$ . It will be shown that over  $N_1(M'_2 - N_2) + M'_2(M'_1 - N_1)$  time slots, we can achieve

$$N_1 \cdot M'_1 \cdot (M'_2 - N_2) \text{ and } N_2 \cdot M'_2 \cdot (M'_1 - N_1)$$

DoF for the two users, respectively. The achievability scheme consists of three phases.

Phase One: This comprises of the initial  $t_1 = N_1(M'_2 - N_2)$  time slots over which T2 remains silent. On the other hand, T1 transmits  $M'_1$  i.i.d. complex Gaussian data symbols intended for R1 in each time slot; and thus, a total of  $N_1 M'_1 (M'_2 - N_2)$  data symbols are sent to R1. We denote these symbols by  $\{u_{1i}(j)\}$ , where  $i \in [1 : M'_1]$ ,  $j \in [1 : t_1]$ , and  $u_{1i}(j)$  are i.i.d. (across  $i$

<sup>3</sup>In this paper, the some variables like  $M'_1$ ,  $M'_2$ ,  $t'$ , etc. have been reused in various sections with different definitions. In each section, follows only the respective definition.

and  $j$ ) according to  $\mathcal{CN}\left(0, \frac{P}{(N_1+N_2)^2}\right)$  distribution. The signal received by R1 is given by

$$Y_1(t) = \tilde{H}_{11}^{[M'_1]}(t) \begin{bmatrix} u_{11}^*(t) & u_{12}^*(t) & \cdots & u_{1M'_1}^*(t) \end{bmatrix}^* + W_1(t), \quad \forall t \in [1 : t_1],$$

where  $\tilde{H}_{ij}^{[x]}(t)$  denotes the matrix obtained from  $H_{ij}(t)$  by retaining only the first  $x$  columns of it. Thus, for a given  $t \in [1 : t_1]$ , R1 observes  $N_1$  linear combinations (LCs) of input data symbols; and these LCs are almost surely linearly independent because  $M'_1 > N_1$  and the Rayleigh-faded channel matrices are almost surely full rank. Since the receiver needs to have one interference-free LC per data symbol to be decoded (provided all LCs are linearly independent of each other) for being able to decode the desired symbols, R1 needs to receive  $M'_1 - N_1$  extra LCs of  $\{u_{1i}(t)\}_i$ ; and as we would see soon, these LCs are present at R2.

R2, over a given time slot of this phase, observes  $N_2$  LCs of data symbols sent to R1. More importantly, any  $M'_1 - N_1 (< N_2)$  LCs out of these would almost surely be linearly independent of each other and also of  $N_1$  LCs observed at that time by R1 (again, because of full-rank property of Rayleigh-faded matrices). Therefore, if R1 is given the signal received at (say) the first  $M'_1 - N_1$  antennas of R2 over the duration of this phase, it can decode the all desired data symbols. We show that this can indeed be accomplished by making use of delayed CSI available to all terminals. With this motivation, let us focus on the signal received by R2 at its first  $M'_1 - N_1$  antennas. At time  $t \in [1 : t_1]$ , it is given by

$$\begin{aligned} Y_{2[1:M'_1-N_1]}(t) &= I_{2[1:M'_1-N_1]}(t) + W_{2[1:M'_1-N_1]}(t), \text{ with} \\ I_{2j}(t) &= \tilde{H}_{2j1}^{[M'_1]}(t) \begin{bmatrix} u_{11}^*(t) & u_{12}^*(t) & \cdots & u_{1M'_1}^*(t) \end{bmatrix}^* \quad \forall j \in [1 : M'_1 - N_1], \end{aligned}$$

where  $\tilde{H}_{2j1}^{[M'_1]}(t)$  is a row vector obtained from another row vector  $H_{2j1}(t)$  by retaining only its first  $M'_1$  entries (see also Fig. 4 for notation). Here,  $\{I_{2j}(t)\}$  is a LC of  $\{u_{1i}(t)\}_i$  whose coefficients are decided by the row vector  $\tilde{H}_{2j1}^{[M'_1]}(t)$ . If suppose R1 knows the LCs  $I_{2[1:M'_1-N_1]}(t)$  as well as the channel matrix  $H_{2[1:M'_1-N_1]}^{[M'_1]}(t)$ , then via channel inversion it can almost surely compute

$$\begin{bmatrix} H_{11}^{[M'_1]}(t) \\ H_{2[1:M'_1-N_1]}^{[M'_1]}(t) \end{bmatrix}^{-1} \begin{bmatrix} Y_1(t) \\ I_{2[1:M'_1-N_1]}(t) \end{bmatrix} = X_{1[1:M'_1]}(t) + (\text{noise terms}) \quad (31)$$

to recover the input symbols  $\{u_{1i}(t)\}_{i=1}^{M'_1}$  (where the noise terms have been ignored since they can not affect the DoF result). Now, R1 by the virtue of delayed CSI, knows the channel matrices

$\{H_{21}(t)\}_{t=1}^{t_1}$  at the end of this phase. Hence, it is sufficient for it to learn the values of the LCs  $\{I_{2j}(t)\}_{j,t}$ . Moreover, at the end of Phase One, T1 knows perfectly the values of LCs  $\{I_{2j}(t)\}_{j,t}$ .

The next phase is analogous to this one during which T2 transmits the  $N_2 M'_2 (M'_1 - N_1)$  number of data symbols to R2. R2, just like the case of R1, does not observe a sufficient number of LCs required to decode the desired data symbols and requires some of the LCs observed by R1. Over the third or the last phase, the transmitters signal such that both the receivers learn the remaining set of LCs that they need to. During the last phase, interference alignment is achieved by making use of the fact that the LCs that are required by the  $i^{th}$  receiver,  $i = 1, 2$ , have already been observed by the  $j^{th}$  receiver,  $j \neq i$ ; and hence, when these LCs are being delivered to the  $i^{th}$  receiver, they do not cause any additional interference to the other receiver. See also Remark 3.

Phase Two: Over this phase of duration  $t_2 = M'_2 (M'_1 - N_1)$  time slots, T1 remains silent, while T2 is transmitting  $M'_2$  i.i.d. complex Gaussian symbols, per time slot, intended for R2. Let these symbols be  $\{u_{2i}(j)\}$ , where  $i \in [1 : M'_2]$ ,  $j \in [1 : t_2]$ , and  $u_{2i}(j)$  are i.i.d. (across  $i$  and  $j$ ) according to  $\mathcal{CN}\left(0, \frac{P}{(N_1 + N_2)^2}\right)$  distribution. Then, if  $t' = t - t_1$ , the signal received by R2 is given by

$$Y_2(t) = \tilde{H}_{22}^{[M'_2]}(t) \begin{bmatrix} u_{21}^*(t') & u_{22}^*(t') & \cdots & u_{2M'_2}^*(t') \end{bmatrix}^* + W_2(t) \quad \forall t \in [t_1 + 1 : t_1 + t_2].$$

Thus, to decode symbols  $\{u_{2i}(t)\}_{i=1}^{M'_2}$ , R2 needs another  $(M'_2 - N_2)$  LCs of them<sup>4</sup>. As argued before, these LCs are present at the first  $M'_2 - N_2$  antennas of R1, where the signal received at time  $t \in [t_1 + 1 : t_1 + t_2]$  is given by

$$\begin{aligned} Y_{1[1:M'_2-N_2]}(t) &= I_{1[1:M'_2-N_2]}(t') + W_{2[1:M'_2-N_2]}(t), \text{ where } t' = t - t_1 \text{ and} \\ I_{1j}(t') &= H_{1j2}^{[M'_2]}(t) \begin{bmatrix} u_{21}^*(t') & u_{22}^*(t') & \cdots & u_{2M'_2}^*(t') \end{bmatrix}^* \quad \forall j \in [1 : M'_2 - N_2]. \end{aligned}$$

As in the case of R1, R2 can decode its desired data symbols, if the values of LCs  $\{I_{ij}(t')\}$  for  $j \in [1 : M'_2 - N_2]$  and  $t' \in [1 : t_2]$  can be conveyed to it because at the end of this phase, R2 will know the channel-dependent coefficients that produce the LCs  $\{I_{ij}(t')\}_{j,t'}$  from the input symbols.

<sup>4</sup>Throughout, whenever we say that a certain receiver needs few more LCs for decoding the desired data symbols, we always mean that these new LCs are linearly independent of each other and also of the LCs this receiver has already observed.

Phase Three: This is the last phase and takes the remaining

$$N_1(M'_2 - N_2) + M'_2(M'_1 - N_1) - t_1 - t_2 = (M'_2 - N_2)(M'_1 - N_1) = t_3$$

time slots. Over this phase, T1 and T2 signal such that the values of  $N_1 t_3$  LCs,  $\{\{I_{2j}(t)\}_{j=1}^{M'_1-N_1}\}_{t=1}^{t_1}$ , are communicated to R1, while those of  $N_2 t_3$  LCs,  $\{\{I_{1i}(t')\}_{i=1}^{M'_2-N_1}\}_{t'=1}^{t_2}$ , are delivered to R2. Once this is accomplished, each receiver gets the one LC per desired data symbols (and all these LCs would be linearly independent) and hence can recover the input symbols.

First, let us partition the total of  $N_1 t_3$  LCs,  $\{I_{2j}(t)\}_{j,t}$ , into  $t_3$  disjoint subsets of cardinality  $N_1$  each<sup>5</sup>. After partitioning, let us relabel these LCs as  $\{I_j^{[2]}(k)\}$ , where  $j \in [1 : N_1]$  and  $k \in [1 : t_3]$ . Similarly, partition  $N_2 t_3$  LCs,  $\{I_{1i}(t')\}_{i,t'}$ , into  $t_3$  subsets of cardinality  $N_2$  each, and after partitioning, let us relabel them as  $\{I_i^{[1]}(k)\}$ , where  $i \in [1 : N_2]$  and  $k \in [1 : t_3]$ . This procedure of partitioning the set of LCs is deterministic and is known to all terminals.

At any time  $t \in [t_1 + t_2 + 1 : t_1 + t_2 + t_3]$  and  $\bar{t} = t - t_1 - t_2$ , T1 and T2 transmit the LCs belonging to the set  $\{I_j^{[2]}(\bar{t})\}_j$  and  $\{I_i^{[1]}(\bar{t})\}_i$ , respectively. Thus, the transmit signals are formed as follows:

$$X_1(t) = \begin{bmatrix} I_1^{[2]}(\bar{t}) \\ I_2^{[2]}(\bar{t}) \\ \vdots \\ I_{N_1}^{[2]}(\bar{t}) \\ 0_{M_1-N_1} \end{bmatrix} \quad \text{and} \quad X_2(t) = \begin{bmatrix} I_1^{[1]}(\bar{t}) \\ I_2^{[1]}(\bar{t}) \\ \vdots \\ I_{N_2}^{[1]}(\bar{t}) \\ 0_{M_2-N_2} \end{bmatrix},$$

where  $0_x$  denotes the all-zero column vector of size  $x$ .

Consider now the decoding procedure at R1. The signal received by it is given by

$$Y_1(t) = H_{11}^{[N_1]}(t) \begin{bmatrix} I_1^{[2]}(\bar{t}) \\ I_2^{[2]}(\bar{t}) \\ \vdots \\ I_{N_1}^{[2]}(\bar{t}) \end{bmatrix} + H_{12}^{[N_2]}(t) \begin{bmatrix} I_1^{[1]}(\bar{t}) \\ I_2^{[1]}(\bar{t}) \\ \vdots \\ I_{N_2}^{[1]}(\bar{t}) \end{bmatrix} + W_1(t), \quad \forall t \in [t_1 + t_2 + 1 : t_1 + t_2 + t_3].$$

<sup>5</sup>Treat here the LC  $I_{2j}(t)$  as a random variable and  $\{I_{2j}(t)\}_{j,t}$  as a set of  $N_1 t_3$  random variables. This step partitions the set  $\{I_{2j}(t)\}_{j,t}$  of random variables into  $t_3$  disjoint subsets.

Since R1 knows LCs  $\{I_i^{[1]}(k)\} \forall i, k$  as well as the channel matrices  $H_{12}(t)$ , it can compute<sup>6</sup>

$$Y'_1(t) = Y_1(t) - H_{12}^{[N_2]}(t) \begin{bmatrix} I_1^{[1]}(\bar{t}) \\ I_2^{[1]}(\bar{t}) \\ \vdots \\ I_{N_2}^{[1]}(\bar{t}) \end{bmatrix} = H_{11}^{[N_1]}(t) \begin{bmatrix} I_1^{[2]}(\bar{t}) \\ I_2^{[2]}(\bar{t}) \\ \vdots \\ I_{N_1}^{[2]}(\bar{t}) \end{bmatrix} + W_1(t) \quad \forall \bar{t} \in [1 : t_3] \quad (32)$$

by subtracting the contribution due to  $\{I_i^{[1]}(\bar{t})\}_{i=1}^{N_2}$ . Subsequently, by inverting channel matrix  $H_{11}^{[N_1]}(t)$  (which can be done with probability 1), it can determine the values of LCs  $\{I_j^{[2]}(\bar{t})\}_{j=1}^{N_1}$ . Hence, as per the arguments developed earlier, R1 can decode all input symbols. The operation of R2 is similar. ■

**Remark 3 (How is interference alignment achieved?):** The signal transmitted by T1 (T2) during this phase, although it is not useful for R2 (R1), does not cause interference to R2 (R1). This is because the symbols, that T1 (T2) is supposed to transmit over this phase, have already been observed by R2 (R1), and hence, their retransmission does not cause any additional interference. In other words, the interference T1 (T2) causes at R2 (R1) over this phase gets ‘aligned’ with the interference it has already caused to R2 (R1) over the first (second) phase. This technique, while it does not produce additional interference at R2 (R1), allows R1 (R2) to learn the remaining set of LCs that it needs to.

## VII. PROOF OF THEOREM 2: CASE A.II

In this case  $N_1 \neq N_2$ , because  $N_1 = N_2 \Rightarrow M_2 < N_2 = N_1$ , which contradicts with the definition of Case A. Thus, for Case A.II, the defining inequalities are

$$M_1, N_1 > N_2 \quad \text{and} \quad N_1 > M_2 \geq N_2.$$

The following lemma shows that the outer-bound is tight.

**Lemma 10:** Under Case A.II, bounds  $L_{\{1,3\}}$  are active and the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is achievable. Moreover, in this case, if

1.  $N_1 \geq M_1$ , then  $L_1$  is active and  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$ ;
2.  $N_1 < M_1$ , then  $L_{\{1,3\}}$  is active and  $\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$ .

<sup>6</sup>R1 knows noisy versions of LCs  $\{I_i^{[1]}(k)\}$ . However, the presence or absence of noise does not alter a DoF result. It is in this sense that we say that R1 knows these LCs.



*Proof:* Note that  $L_2$  is implied by  $L_3$ , because for this case

$$L_2 \equiv d_1 + d_2 \leq \min(N_1, M_1 + M_2) \text{ and } L_3 \equiv d_1 + d_2 \leq \min(N_1, M_1)$$

with  $\min(N_1, M_1 + M_2) \geq \min(N_1, M_1)$ , which proves the first part of the lemma. For an easy reference, we restate bound  $L_1$ :

$$L_1 \equiv \frac{d_1}{\min(M_1, N_1 + N_2)} + \frac{d_2}{N_2} \leq 1.$$

Again, depending upon whether  $L_1$  is active and/or  $L_3$  is active, we have two sub-cases:

- Case A.II.1:  $N_1 \geq M_1 \Rightarrow N_1 \geq M_1 > N_2$  and  $N_1 > M_2 \geq N_2$ :

Here,  $L_1 \equiv \frac{d_1}{M_1} + \frac{d_2}{N_2} \leq 1$  and  $L_3 \equiv d_1 + d_2 \leq M_1$ . Since  $N_2 < M_1$ ,  $L_1$  implies  $L_3$ , and thus  $L_1$  is active. It can be verified that in this case, the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  with  $L_1$  active coincides with  $\mathbf{D}^{\text{no-CSI}}$ , which yields us  $\mathbf{D}^{\text{no-CSI}} = \mathbf{D}^{\text{d-CSI}}$  and hence the lemma.

- Case A.II.2:  $N_1 < M_1 \Rightarrow M_1 > N_1 > M_2 \geq N_2$ :

Here, both the bounds  $L_{\{1,3\}}$  are strictly active. The typical shape of the outer-bound is as shown in Fig. 2(e) from which we see that the achievability of point  $P_{1,3}$  (the point of intersection of bounds  $L_1$  and  $L_3$ ) is sufficient to establish the achievability of outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$ . In the following, we develop a scheme to achieve point  $P_{1,3}$ .

An achievability scheme for point  $P_{1,3}$ :

Here,

$$P_{1,3} \equiv \left( \frac{M'_1(N_1 - N_2)}{M'_1 - N_2}, \frac{N_2(M'_1 - N_1)}{M'_1 - N_2} \right),$$

where  $M'_1 = \min(M_1, N_1 + N_2)$ . In the following achievability scheme, we make use of only  $M'_1$  antennas at T1, and thus it is assumed henceforth in this section that  $M_1 = M'_1$ . The basic idea behind the following achievability scheme, although it consists of two phases as opposed to three, is similar to the one presented in the last section, and therefore, to avoid repetition, we just present an outline.

It will be shown that over  $M_1 - N_2$  time slots,  $M_1(N_1 - N_2)$  and  $N_2(M_1 - N_1)$  DoF can be achieved for the two users, respectively (recall the assumption of  $M_1 = M'_1$ ).

Phase One: Over this phase, which comprises of  $t_1 = (N_1 - N_2)$  time slots, T1 transmits  $M_1(N_1 - N_2)$  i.i.d. complex Gaussian data symbols  $\{u_{1i}(j)\}$ , where  $i \in [1 : M_1]$  and  $j \in [1 : N_1 - N_2]$ , intended for R1 whereas T2 remains silent. At a given time  $t \in [1 : N_1 - N_2]$ , R1 receives  $N_1$  LCs of input symbols  $\{u_{1i}(t)\}_{i=1}^{M_1}$ ; and therefore, to decode them, it needs extra

$M_1 - N_1$  LCs of them, which, as one may expect, are observed by R2 at the first  $M_1 - N_1$  antennas (recall  $M'_1 = M_1 \leq N_1 + N_2$ ). Mathematically, the LCs observed by R2 at its  $j^{th}$  antenna is given by

$$I_{2j}(t) = H_{2j1}(t) \begin{bmatrix} u_{11}^*(t) & u_{12}^*(t) & \cdots & u_{1M_1}^*(t) \end{bmatrix}^* \quad \forall t \in [1 : t_1],$$

and if R1 knows the values of  $I_{2[1:M_1-N_1]}(t)$  and the channel matrix  $H_{21}(t)$ , it can decode the data symbols  $\{u_{1i}(t)\}_i$  (c.f. equation (31)). Moreover, R1 knows the all channel vectors  $\{H_{2[1:M_1-N_1]1}(t)\}_{t=1}^{t_1}$  at the end of this phase.

Since T1 has already transmitted the required number of data symbols intended for R1, its goal over the next phase is to provide R1 the values of LCs  $\{I_{2j}(t)\}_{j,t}$ , which it knows at the end of this phase. As opposed to it, T2 is yet to transmit and thus, sends all  $N_2(M_1 - N_1)$  number of data symbols intended for R2 over the next phase.

Phase Two: This is the last phase and occupies

$$(M_1 - N_2) - t_1 = (M_1 - N_2) - (N_1 - N_2) = (M_1 - N_1) = t_2$$

time slots. For a given  $t \in [t_1 + 1 : t_1 + t_2]$  and  $t' = t - t_1$ , the transmit signals are formed as follows:

$$X_1(t) = \begin{bmatrix} I_{2t'}(1) \\ I_{2t'}(2) \\ \vdots \\ I_{2t'}(t_1) \\ 0_{M_1-t_1} \end{bmatrix} \quad \text{and} \quad X_2(t) = \begin{bmatrix} u_{21}(t') \\ u_{22}(t') \\ \vdots \\ u_{2N_2}(t') \\ 0_{M_2-N_2} \end{bmatrix},$$

where  $\{u_{2j}(t')\}$  with  $j \in [1 : N_2]$  and  $t' \in [1 : t_2]$  are i.i.d. complex Gaussian data symbols intended for R2. Since  $t_1 = N_1 - N_2$ , exactly  $N_1$  transmit antennas are in use (i.e., they are transmitting a non-zero signal while the rest of set to zero) at any given time during this phase.

Since R2 knows interfering symbols  $\{I_{2t'}(j)\}$  for  $\forall j, t'$ , it can subtract their contribution from its received signal (c.f. equation (32)) and then via channel inversion can recover the desired data symbols (c.f. (31)).

Further, since exactly  $N_1$  antennas are in use during this phase, R1 can zero-force the interference coming from T2 to recover the signal sent by T1. Specifically, it can simply ignore the last  $M_1 - t_1$  and  $M_2 - N_2$  antennas of T1 and T2, respectively (because these are set to zero),

and then can almost surely invert the square channel matrix  $\begin{bmatrix} H_{11}^{[t_1]}(t) \\ H_{12}^{[N_2]}(t) \end{bmatrix}$  to itself from the first  $t_1$  and the first  $N_2$  transmit antennas of T1 and T2, respectively, to recover the LCs  $I_{2t'}([1 : N_2])$  (and also  $\{u_{2[1:N_2]}(t')\}$ ). Thus, at the end of this phase, it gets one LC per desired data symbol and hence can decode all of them. ■

### VIII. PROOF OF THEOREM 2: CASES B.0 AND B.I

For Case B.0, we have the following lemma.

**Lemma 11:** Under Case B.0,  $\mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$ , which implies that bound  $L_3$  is active.

*Proof:* Under Case B.0, the inequality  $N_1 = N_2 > M_2$  holds. The lemma now follows from Remark 17 of [16]. ■

Henceforth, in this section, we study Case B.I, for which the defining inequality is

$$N_1 \geq M_1 > N_2 > M_2$$

and we have the following lemma.

**Lemma 12:** Under Case B.I, bound  $L_1$  is active and the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is achievable.

*Proof:* Here, the three outer-bounds are given as

$$L_1 \equiv \frac{d_1}{M_1} + \frac{d_2}{N_2} \leq 1, \quad L_2 \equiv d_1 + d_2 \leq \min(N_1, M_1 + M_2), \quad \text{and} \quad L_3 \equiv d_1 + d_2 \leq M_1.$$

It can be easily verified that  $L_3$  implies  $L_2$  and since  $N_2 < M_1$ ,  $L_1$  implies  $L_3$ . Hence, bound  $L_1$  is active.

The typical shape of the outer-bound is shown in Fig. 3(a). It is clear that if the point

$$P_{o2,1} \equiv \left( \frac{M_1}{N_2}(N_2 - M_2), M_2 \right),$$

which is the point of intersection of the single-user bound on  $d_2$  and the bound  $L_1$ , is known to be achievable, then the entire outer-bound can be achieved via time sharing.

A scheme to achieve point  $P_{o2,1}$ :

Over  $N_2$  time slots, we show the achievability of the DoF tuple  $(M_1(N_2 - M_2), N_2 M_2)$ . In this scheme, R1 makes use of only  $M_1$  receive antennas (even when more are available), and thus henceforth, in this section, it is assumed that  $N_1 = M_1$ . In this scheme, T2 at all times needs to transmit  $M_2$  i.i.d. complex Gaussian data symbols intended for R2.

Define unitary matrices  $U_{12}(t)$  and  $U_{22}(t)$  such that last  $N_1 - M_2$  and  $N_2 - M_2$  rows of  $U_{12}(t)H_{12}(t)$  and  $U_{22}(t)H_{22}(t)$ , respectively, consist only of zeros. Note that these unitary matrices are functions of the channel matrices, and hence, any terminal can compute these as soon as it knows the channel matrices. Set  $Y'_i(t) = U_{i2}(t)Y_i(t)$ ,  $i = 1, 2$ . At each time, the  $i^{th}$  receiver computes  $Y'_i(t)$  and uses the signal  $Y'_i(t)$ , instead of  $Y_i(t)$ , for all further decoding purposes. Note that the transmit signal  $X_2(t)$  affects only the first  $M_2$  entries of  $Y'_i(t) \forall i, t$ .

Since T2 is always transmitting  $M_2$  data symbols, R1, along the first  $M_2$  entries of  $Y'_1(t)$ , receives interference. In the following scheme, R1 simply discards the first  $M_2$  entries of  $Y'_1(t)$  and retains only the remaining  $M_1 - M_2$  entries (recall  $N_1 = M_1$ ).

The achievability scheme, as described below, consists of two phases.

Phase One: It lasts for  $t_1 = N_2 - M_2$  time slots. At time  $t \in [1 : t_1]$ , T1 transmits  $M_1$  i.i.d. complex Gaussian data symbols,  $\{u_{1i}(t)\}_{i=1}^{M_1}$ , intended for R1. For a given  $t \in [1 : t_1]$ , R1 receives  $M_1 - M_2$  LCs of the input data symbols sent to it by T1, and requires another  $M_2$  LCs of them.

Consider now the case of R2. It receives the desired data symbols along the first  $M_2$  entries of  $Y'_2(t)$ . However, since  $M_1 > N_2 > M_2$ , these entries are corrupted by interference due to the signal coming from T1. More specifically, the signal received by R2 during this phase is given by

$$Y'_2(t) = U_{22}(t)H_{22}(t) \begin{bmatrix} u_{21}^*(t) & u_{22}^*(t) & \cdots & u_{2M_2}^*(t) \end{bmatrix}^* + I_2(t) + W_2(t)$$

where

$$I_2(t) = U_{22}(t)H_{21}(t) \begin{bmatrix} u_{11}^*(t) & u_{12}^*(t) & \cdots & u_{1M_1}^*(t) \end{bmatrix}^*. \quad (33)$$

Hence, if R2 knows the interfering symbols  $I_{2[1:M_2]}(t)$ , it can compute

$$Y'_{2[1:M_2]}(t) - I_{2[1:M_2]}(t) = U_{22}^{[M_2]}(t)H_{22}(t) \begin{bmatrix} u_{21}^*(t) & u_{22}^*(t) & \cdots & u_{2M_2}^*(t) \end{bmatrix}^* + W_2(t) \quad (34)$$

and then by inverting the square matrix  $U_{22}^{[M_2]}(t)H_{22}(t)$ , it can recover the desired symbols. In other words, R2 needs to know the interfering symbols  $\{I_{2j}(t)\}_{j=1}^{M_2}$ . Moreover, since the interfering symbols  $\{I_{2j}(t)\}_{j=1}^{M_2}$  are just the LCs of  $\{u_{1i}(t)\}_{i=1}^{M_1}$  and there are precisely  $M_2$  of these symbols, their knowledge would almost surely enable R1 to decode its desired symbols. Hence, it is beneficial to convey these interfering symbols to both the receivers, which is the goal of the next phase. Note that at the end of this phase, T1 knows  $\{\{I_{2j}(t)\}_{j=1}^{M_2}\}_{t=1}^{t_1}$  perfectly,

while both the receivers know the channel-dependent coefficients that produce these interfering symbols (or the LCs).

Phase Two: This last phase takes the remaining  $M_2$  time slots. Here, T1 transmits the interfering symbols  $\{I_{2j}(t)\}_{j,t}$  so that both R1 and R2 can determine these, while R2 continues to receive  $M_2$  new data symbols from T2 in each slot. For a given  $t \in [t_1 + 1 : t_1 + M_2]$  and  $t' = t - t_1$ , the transmit signals are formed as follows:

$$X_1(t) = \begin{bmatrix} I_{2t'}(1) \\ I_{2t'}(2) \\ \vdots \\ I_{2t'}(t_1) \\ 0_{M_1-t_1} \end{bmatrix} \quad \text{and} \quad X_2(t) = \begin{bmatrix} u_{21}(t) \\ u_{22}(t) \\ \vdots \\ u_{2M_2}(t) \end{bmatrix}.$$

Thus, during this phase, a total  $t_1 + M_2 = N_2$  transmit antennas are in use at any given time (i.e., the rest of the transmit antennas are sending nothing). Since  $N_1 > N_2$ , both the receivers can recover the transmit signals almost surely via channel inversion, as explained in the previous section.

Thus, R2 can learn all new desired data symbols sent during this phase as well as all the interfering symbols  $\{I_{2j}(t)\}_{j,t}$ , whose knowledge would enable it to decode the desired data symbols sent over Phase One. At the same time, R1 gets one interference-free LCs per desired symbol and hence can decode them. ■

**Remark 4 (On interference alignment):** In the above achievability scheme, interference alignment is achieved by making T1 transmit the interfering symbols over Phase Two, which have already caused interference to R2. This serves three purposes: First, it allows R1 to learn the LCs it has lost due to interference. Second, it does not cause any additional interference at R2. Third, it gives R2 an opportunity to learn the interference it suffered over Phase One and thereby enabling it to decode the desired data symbols sent over the previous phase.

## IX. PROOF OF THEOREM 2: CASE B.II

First note from Lemma 20 that Condition 1 can be expressed in an equivalent form as given by

$$M_1 > M'_1 > N_1 > N_2 > M_2 > m$$

where  $M'_1 = \min(M_1, N_1 + N_2 - M_2)$  and  $m = N_2 \frac{M'_1 - N_1}{M'_1 - N_2}$ . Thus, the defining inequality for this case is

$$M_1 > N_1 > N_2 > M_2 \text{ and either } M_1 = M'_1 \text{ or } M_2 \leq m.$$

Then we have the following lemma.

**Lemma 13:** Under Case B.II, bounds  $L_{\{1,3\}}$  are active and, moreover, if

$$N_2 \cdot \frac{M_1 - N_1}{M_1 - N_2} \geq M_2, \quad (35)$$

bound  $L_3$  is active. Further, the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  is achievable. More specifically, if

1.  $M_2 \leq m$ ,  $L_3$  is active and  $\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$ ,
2.  $M_1 = M'_1$  and  $M_2 > m$ ,  $L_{\{1,3\}}$  is active and  $\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$ .

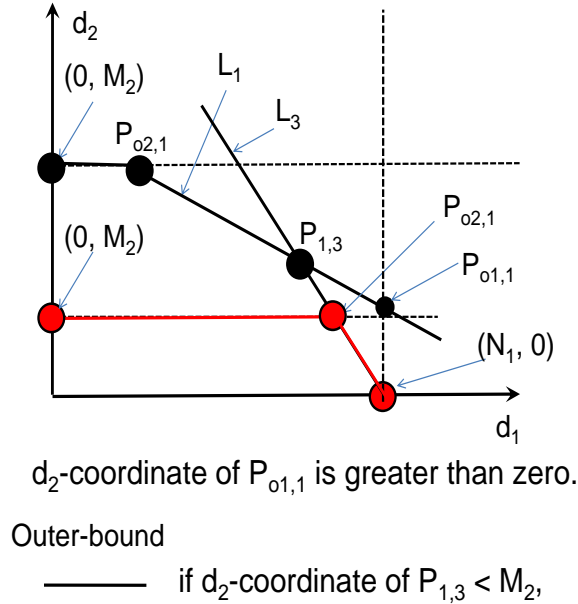


Fig. 5. Two Possible Shapes of the Outer-Bound: Case B.II

*Proof:* A simple evaluation shows that  $L_3$  implies  $L_2$  (see the proof presented for Case B.I). Since the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  does change even if  $M_1$  increases beyond  $N_1 + N_2$  and since all the achievability schemes developed here make use of at most  $N_1 + N_2$  antennas of T1, it may be assumed, without loss of generality, that  $M_1 = \min(M_1, N_1 + N_2)$ . Then, the bounds  $L_1$  and  $L_3$ , which are active, are given by

$$L_1 \equiv \frac{d_1}{M_1} + \frac{d_2}{N_2} \leq 1 \text{ and } L_3 \equiv d_1 + d_2 \leq N_1.$$

The straight lines corresponding to  $L_1$  and  $L_3$  have unequal slopes, and hence intersect. From Fig. 5, one may observe that if the  $d_2$ -coordinate of point  $P_{1,3}$  (the point of intersection of  $L_1$  and  $L_3$ ), is greater than or equal to  $M_2$ , only bound  $L_3$  is active. Since the  $d_2$ -coordinate of  $P_{1,3}$  is equal to  $N_2 \frac{M_1 - N_1}{M_1 - N_2}$ , bound  $L_3$  is active when the inequality (35) is true.

We now proceed to the last part of the lemma. Under Case B.II, where the inequality  $M_1 > N_1 > N_2 > M_2$  always holds, Condition 1 can not hold if either of the following two conditions are true: First,  $M_2 \leq m$ , and second,  $\{M_1 \leq N_1 + N_2 - M_2 \text{ and } M_2 > m\}$ . This yields the following two subcases:

- Case B.II.1:  $M_2 \leq m \Rightarrow M_1 > N_1 > N_2 > M_2$  and  $M_2 \leq m$ :

In this case, we have

$$M_2 \leq m \leq N_2 \frac{M_1 - N_1}{M_1 - N_2},$$

where the last inequality is true because for a given  $N_1$  and  $N_2$ ,  $\frac{M_1 - N_1}{M_1 - N_2}$ , as a function of  $M_1$ , is increasing and because  $M_1 \geq M'_1$ . This implies that  $L_3$  is active and  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}} = \mathbf{D}^{\text{p-CSI}}$ .

- Case B.II.2:  $\{M_1 \leq N_1 + N_2 - M_2 \text{ and } M_2 > m\} \Rightarrow M'_1 = M_1 > N_1 > N_2 > M_2 > m$ :

Here, bounds  $L_{\{1,3\}}$  are (strictly) active.

For each of the above cases, we develop an achievability scheme to exhaust the outer-bound.

*A. Case B.II.1:  $M_1 > N_1 > N_2 > M_2$  and  $M_2 \leq N_2 \frac{M'_1 - N_1}{M'_1 - N_2}$*

It is sufficient to prove the achievability of  $P_{o2,3} \equiv (N_1 - M_2, M_2)$ ; see Fig. 3(b).

A scheme to achieve point  $P_{o2,3}$ :

It will be shown that over  $M'_1$  time slots,  $M'_1(N_1 - M_2)$  and  $M'_1 M_2$  DoF can be achieved for the two users, respectively. T2 needs to send  $M_2$  i.i.d. complex Gaussian data symbols over each time slot. As done in the previous section, define  $Y'_1(t)$  and  $Y'_2(t)$  such that the transmit signal  $X_2(t)$  affects only the first  $M_2$  entries of these. Again, the  $i^{\text{th}}$  receiver uses  $Y'_i(t)$  for all decoding purposes. Further, at each time slot, R1 ignores the first  $M_2$  entries of  $Y'_i(t)$ , which carry interference. The transmission scheme consists of two phases.

Phase One: It consists of the first  $t_1 = N_1 - M_2$  time slots. At time  $t \in [1 : t_1]$ , T1 sends  $M'_1$  i.i.d. complex Gaussian data symbols,  $\{u_{1i}(t)\}_{i=1}^{M'_1}$ , intended for R1. Now, R1, after ignoring the first  $M_2$  entries of  $Y'_1(t)$ , can be said to observe  $N_1 - M_2$  LCs of these symbols at a given time  $t \in [1 : t_1]$ , and thus requires another  $M'_1 - (N_1 - M_2)$  LCs of them. As done in the previous

sections, these LCs required by R1 are present at the antennas of R2. With this motivation, consider the signal received by R2 at time  $t \in [1 : t_1]$ :

$$Y'_2(t) = U_{22}(t)H_{22}(t) \begin{bmatrix} u_{21}^*(t) & \cdots & u_{2M_2}^*(t) \end{bmatrix}^* + I_2(t) + W_2(t),$$

where

$$I_2(t) = U_{22}(t)H_{21}^{[M'_1]}(t) \begin{bmatrix} u_{11}^*(t) & \cdots & u_{1M'_1}^*(t) \end{bmatrix}^*.$$

It should be clear that R1 can almost surely decode the input symbols  $\{u_{1i}(t)\}_i$ , if it knows the values of the LCs  $I_{2[1:M_2+(M'_1-N_1)]}(t)$ .

Consider now the case of R2, which receives its useful signal along the first  $M_2$  entries of  $Y'_2(t)$ . Since these entries are interfered by the LCs  $I_{2[1:M_2]}(t)$ , R2 needs to know these values so that it can subtract them from  $Y'_2(t)$  to recover the desired data symbols (cf. (34)). Hence, the LCs  $I_{2[1:M_2]}(t)$  are required by both the receivers. Further, since the signal sent by T2 affects only the first  $M_2$  entries of  $Y'_2(t)$ , R2 knows the values of LCs  $I_{2[M_2+1:M_2+(M'_1-N_1)]}(t)$ , which are required by R1.

Thus, over the next phase, T1 needs to signal such that both the receivers get LCs  $\{I_{2[1:M_2]}(t)\}_{t=1}^{t_1}$ , and R1 additionally receives  $I_{2[M_2+1:M_2+(M'_1-N_1)]}(t)$ , while at the same time, R2 is able to receive  $M_2$  new data symbols in each time slot from T2.

Before starting the description of the next phase, we introduce some terminology. An element of set  $\mathcal{S} = \{I_{2[1:M_2+(M'_1-N_1)]}(t)\}_{t=1}^{t_1}$  is referred to in the sequel as the ‘interfering symbol’. This set can be partitioned into two disjoint subsets:  $\mathcal{S}_{\text{R2-known}} = \{I_{2[M_2+1:M_2+(M'_1-N_1)]}(t)\}_{t=1}^{t_1}$ , which consists of interfering symbols that are known to R2 but are required by R1; and  $\mathcal{S}_{\text{R2-unknown}} = \{I_{2[1:M_2]}(t)\}_{t=1}^{t_1}$ , which consists of interfering symbols that are required by both the receivers (unknown to R2, as well).

Phase Two: It occupies the remaining  $M'_1 - t_1 = t_2 = (M'_1 + M_2 - N_1)$  time slots. Here, over each slot, T1 sends some of the interfering symbols. Recall that all the interfering symbols are required by R1, at which only  $N_1 - M_2$  dimensions are available for its useful signal. Hence, T1, at no time during this phase, should transmit more than  $N_1 - M_2$  interfering symbols. Moreover, some of these interfering symbols, in particular, those belonging to the set  $\mathcal{S}_{\text{R2-unknown}}$  are required by R2 as well, which implies that T1 should never transmit more than  $N_2 - M_2$  elements of the set  $\mathcal{S}_{\text{R2-unknown}}$  (recall, R2 needs to decode  $M_2$  data symbols per time slot). Nonetheless,



under these constraints, T1 should be able to transmit every element of the set  $\mathcal{S}$  (at least once). Hence, the transmit signal of T1 needs to be designed to meet the above objectives. To this end, the following lemma helps.

**Lemma 14:** There exists a fixed, deterministic partition  $\mathcal{P}$  of the set  $\mathcal{S}$  into  $t_2$  disjoint subsets  $\{\mathcal{S}(j)\}_{j=1}^{t_2}$  of cardinality  $(N_1 - M_2)$  each such that no subset contains more than  $(N_2 - M_2)$  elements of set  $\mathcal{S}_{\text{R2-unknown}}$ .

*Proof:* This lemma rests on the fact that  $M_2 \leq m$ . See Appendix B for the proof. ■

Consider a partition  $\mathcal{P}$  given by the above lemma. Let us denote the elements of subset  $\mathcal{S}(j)$  by  $\{I_i^{[2]}(j)\}_{i=1}^{N_1-M_2}$ . At time  $t \in [t_1 + 1 : t_1 + t_2]$ , T1 transmits elements of set  $\mathcal{S}(t - t_1)$  using the first  $N_1 - M_2$  of its antennas in a one-to-one fashion, while T2 sends  $M_2$  new data symbols. Thus, for a given  $t \in [t_1 + 1 : t_1 + t_2]$  and  $t' = t - t_1$ , T1 and T2 construct their signals as follows:

$$\begin{aligned} X_{1i}(t) &= I_i^{[2]}(t') \quad \forall i \in [1 : N_1 - M_2], \quad X_{1i}(t) = 0 \quad \forall i \in [N_1 - M_2 + 1 : M_1] \quad \text{and} \\ X_{2j}(t) &= u_{2j}(t) \quad \forall j \in [1 : M_2]. \end{aligned} \quad (36)$$

Since the cardinality of  $\mathcal{S}(t')$  can never exceed  $N_1 - M_2$ , exactly  $N_1$  transmit antennas are in use at any given time during this phase. Hence, R1 can zero-force the interference coming from T2 and recover the signal sent by T1. Thus, at the end of this phase, R1 gets one interference-free LC per desired data symbol, and hence can decode all desired data symbols.

Next, consider R2. For no  $t' \in [1 : t_2]$ , a subset  $\mathcal{S}(t')$  contains more than  $N_2 - M_2$  elements of the set  $\mathcal{S}_{\text{R2-unknown}}$ , which implies that T1 and T2 together are transmitting at most  $N_2$  symbols that are unknown to R2 (the rest are known to it). Hence, R2 can subtract the contribution due to the known interfering symbols (cf. (32)), and then via channel inversion can recover the desired data symbols as well as the unknown interfering symbols. Once it knows the unknown interfering symbols, it can subtract the contribution due to these from the signal it has received over Phase One and then again via channel inversion can recover the desired data symbols sent to it over Phase One.

**B. Case B.II.2:**  $N_1 + N_2 - M_2 \geq M_1 > N_1 > N_2 > M_2 > N_2 \frac{M_1 - N_1}{M_1 - N_2}$

In this case, bounds  $L_{\{1,3\}}$  are active and a typical shape of the outer-bound is as shown in Fig. 3(c). We need to establish the achievability of points  $P_{o2,1}$  and  $P_{1,3}$ . We develop two separate

achievability schemes for these two points.

1) An achievability scheme for point  $P_{o2,1}$ :

Here, point  $P_{o2,1}$  is given by

$$P_{o2,1} \equiv \left( \frac{M_1}{N_2}(N_2 - M_2), M_2 \right).$$

Recall that in the context of Case B.I, we developed an achievability scheme for point  $P_{o2,1}$  in the last section. Here, we modify this scheme such that it works for the present case. In the following, we will just point out the important differences. In the scheme of the present section, R1 makes use of all its antennas, unlike the one of the last section where it uses only  $M_1$  of its antennas even if  $N_1 \geq M_1$  (which is the case in the last section).

The objective is to achieve a DoF tuple  $(M_1(N_2 - M_2), M_2 N_2)$  over  $N_2$  time slots. Recall T2 needs to transmit  $M_2$  i.i.d. data symbols in each time slot.

Phase One: It again consists of  $t_1 = N_2 - M_2$  time slots. T1 transmits  $M_1$  i.i.d. complex data symbols in each time slot. R1 experiences interference at  $M_2$  of its receive dimensions and thus observes  $N_1 - M_2$  interference-free LCs per time slot, which also implies that it needs another  $M_1 + M_2 - N_1$  LCs. Note that up to this point, the scheme is exactly identical to the one of the last section, except that we let R1 use all  $N_1$  antennas and have accounted for the fact that  $M_1$  is now greater than  $N_1$ .

Now, R2 receives its useful signal along the first  $M_2$  entries of  $Y_2'(t)$  (which is obtained from  $Y_2(t)$  after a suitable unitary transformation). Hence, as argued in the last section, R2 can decode its desired signal if it knows the interfering symbols  $I_{2[1:M_2]}(t)$  (see equation (33) of the last section).

On the other hand, R1 can decode the desired data symbols sent at time  $t$  if it knows the values of LCs  $I_{2[1:M_1+M_2-N_1]}(t)$  ( $M_1 + M_2 - N_1 \leq N_2$ ). Recall that in the scheme of the previous section, it was sufficient for R1 to know just the values of  $I_{2[1:M_2]}(t)$ ; whereas in the present case it needs to learn more – a difference due to  $M_1 > N_1$  here, whereas in the last section, we had  $M_1 \leq N_1$ . Also note that since  $X_2(t)$  affects only the first  $M_2$  entries of  $Y_2'(t)$ , the values of LCs  $I_{2[M_2+1:M_1+M_2-N_1]}(t)$  are known to R2.

Thus over the next phase, T1 needs to signal such that the values of LCs  $\{I_{2[1:M_2]}(t)\}_{t=1}^{t_1}$  are conveyed to both the receivers, and additionally, those of  $\{I_{2[M_2+1:M_1+M_2-N_1]}(t)\}_{t=1}^{t_1}$  are delivered to R1, and simultaneously R2 is able to decode  $M_2$  new i.i.d. data symbols coming from T2 in

each slot.

Define  $\mathcal{S}_{\text{R2-known}} = \{ I_{2[M_2+1:M_2+M_1-N_1]}(t) \}_{t=1}^{t_1}$ ,  $\mathcal{S}_{\text{R2-unknown}} = \{ I_{2[1:M_2]}(t) \}_{t=1}^{t_1}$ , and  $\mathcal{S} = \mathcal{S}_{\text{R2-unknown}} \cup \mathcal{S}_{\text{R2-known}}$  – the elements of which are called the interfering symbols. For a set  $\mathcal{S}$ ,  $|\mathcal{S}|$  denotes its cardinality.

Partition the set  $\mathcal{S}_{\text{R2-known}}$  into  $t_2 = N_2 - t_1 = M_2$  disjoint subsets  $\mathcal{S}_{\text{R2-known}}(j)$ , where  $j \in [1 : t_2]$ , such that no subset has cardinality more than  $N_1 - N_2$ , and as per the following lemma, this is feasible.

**Lemma 15:**  $|\mathcal{S}_{\text{R2-known}}| = (M_1 - N_1)(N_2 - M_2) \leq t_2(N_1 - N_2)$ .

*Proof:* The first equality follows from the definition. Suppose the second inequality is not true. Then we have

$$(M_1 - N_1)(N_2 - M_2) > M_2(N_1 - N_2) \Rightarrow \frac{N_2 - M_2}{M_2} > \frac{N_1 - N_2}{M_1 - N_1} \Rightarrow \frac{N_2}{M_2} > \frac{M_1 - N_2}{M_1 - N_1},$$

which contradicts with the defining inequality of Case B.II.2. Hence, the second inequality is true. ■

Let the elements of  $\mathcal{S}_{\text{R2-known}}(j)$  be  $I_i^{[2]}(j)$ ,  $i \in [1 : k_j]$ , where  $k_j$  is some integer such that  $1 \leq k_j \leq (N_1 - N_2) \forall j$ .

Phase Two: T1 transmits interfering symbols over this phase. At time  $t \in [t_1 + 1 : t_1 + t_2]$  with  $t' = t - t_1$ , the transmit signals are formed as follows:

$$\begin{aligned} X_1(t) &= \left[ I_{2t'}^*(1) \quad \cdots \quad I_{2t'}^*(N_2 - M_2) \quad I_1^{[2]}(t') \quad \cdots \quad I_{k_{t'}}^{[2]}(t') \quad 0_{M_1 - (N_2 - M_2 + k_{t'})} \right]^* \text{ and} \\ X_2(t) &= \left[ u_{21}^*(t) \quad \cdots \quad u_{2M_2}^*(t) \right]. \end{aligned}$$

Since  $(N_2 - M_2) + k_{t'} + M_2 \leq N_1 \forall t'$ , not more than  $N_1$  transmit antennas are in use at any time during this phase. Hence, R1 can recover the interfering symbols after zero-forcing the interference coming from T2. Further, since R2 knows all elements of set  $\mathcal{S}_{\text{R2-known}}$ , exactly  $N_2$  symbols unknown to R2 are being transmitted at any given time. Hence, R2, after subtracting the contribution due to known interfering symbols, can recover the desired data symbols as well as the unknown interfering symbols.

Thus, at the end of each phase, each receiver get one interference-free LC per desired data symbol and hence each of them can decode the desired symbols.

## 2) An achievability Scheme for Point $P_{1,3}$ :

We have

$$P_{1,3} \equiv \left( \frac{M_1(N_1 - N_2)}{M_1 - N_2}, \frac{N_2(M_1 - N_1)}{M_1 - N_2} \right).$$

We now describe below a generic IA-based achievability scheme,  $\mathcal{IA}$ -Scheme, which is defined in terms of certain set of parameters. This scheme with appropriate choice for parameters can be used to achieve point  $P_{1,3}$  above. Moreover, it is useful in the next section as well.

Let  $\mathbb{N}$  and  $\mathbb{N}_0$  be the sets of positive and non-negative integers, respectively. Consider the achievability scheme  $\mathcal{IA}$ -Scheme  $\left\{ (M_1, M_2, N_1, N_2), (d_1^*, d_2^*), (T, t_1, t_2), \{m_t\}_{t=1}^{t_1}, \{n_t\}_{t=1}^{t_1} \right\}$ , where  $d_1^*, d_2^*, T, t_1, t_2 \in \mathbb{N}$ ;  $m_t, n_t \in \mathbb{N}_0 \forall t \in [1 : t_1]$ ;  $t_1 < T, t_2 = T - t_1$ ;  $N_1 \leq m_t \leq M_1, n_t \leq M_2$ ,  $m_t + n_t \leq N_1 + N_2 \forall t \in [1 : t_1]$ ; and  $\sum_{t=1}^{t_1} m_t = d_1^*$ . We will design this coding scheme such that over  $T$  time slots,  $d_1^*$  and  $d_2^*$  DoF can be achieved respectively for the two users, provided the parameters  $t_1, \{m_t\}, \{n_t\}$  are chosen appropriately. Thus, using this scheme, it is possible to achieve the DoF pair  $\left( \frac{d_1^*}{T}, \frac{d_2^*}{T} \right)$ .

The scheme takes a total of  $T$  time slots and consists of two phases.

Phase 1: This phase constitutes the initial  $t_1$  time slots. At time  $t \in [1 : t_1]$ , T1 transmits  $m_t$  i.i.d. complex Gaussian data symbols,  $\{u_{1j}(t)\}_{j=1}^{m_t}$ , intended for R1; whereas T2 transmits  $n_t$  i.i.d. complex Gaussian data symbols,  $\{u_{2j}(t)\}_{j=1}^{n_t}$ , intended for R2. Note that this is feasible since by definition  $m_t \leq M_1$  and  $n_t \leq M_2$ .

Consider now the scenario at the receivers. It may be assumed, without loss of generality, that the signal transmitted by T2 at time  $t \in [1 : t_1]$  affects only the first  $n_t$  entries of  $Y_1(t)$  and  $Y_2(t)$  (cf. the discussion in Section VIII on the achievability of point  $P_{o2,1}$ ). Thus, R1 observes  $N_1 - n_t$  LCs of input symbols  $\{u_{1j}(t)\}_{j=1}^{m_t}$  without any interference, and therefore, needs another  $m_t - (N_1 - n_t)$  LCs to decode these symbols. The additional LCs required by R1 are present at (some of the) antennas of R2 because  $m_t \leq N_1 + N_2 - n_t \forall t \in [1 : t_1]$ . More precisely, the signal received by R2 at time  $t$  during this phase is given by

$$\begin{aligned} Y_2(t) &= H_{22}^{[n_t]}(t) \begin{bmatrix} u_{21}^*(t) & \cdots & u_{2n_t}^*(t) \end{bmatrix} + I_2(t) + W_2(t), \text{ where} \\ I_2(t) &= H_{21}^{[m_t]}(t) \begin{bmatrix} u_{11}^*(t) & \cdots & u_{1m_t}^*(t) \end{bmatrix}^*. \end{aligned}$$

Then R1 can decode the desired data symbols  $\{u_{1j}(t)\}_{j=1}^{m_t}$ , if it knows the values of LCs  $\{I_{2[1:n_t+(m_t-N_1)]}(t)\}$ . Now, R2, at time  $t$ , receives the desired symbols along the first  $n_t$  entries of  $Y_2(t)$ , which experience interference due to  $I_{2[1:n_t]}(t)$ . Thus, R2 can decode its desired data

symbols, provided it knows the values of  $I_{2[1:n_t]}(t)$ . It must be noted that R2 knows the values of  $\{I_{2[n_t+1:n_t+(m_t-N_1)]}(t)\}$ .

Thus, over the next phase, we have to accomplish the following objectives: *a)* communicate the values of LCs  $\{I_{2[1:n_t]}(t)\}_{t=1}^{t_1}$  to both the receivers, *b)* additionally, deliver those of  $\{I_{2[n_t+1:n_t+(m_t-N_1)]}(t)\}$  to R1, and *c)* transmit  $n' = d_2^* - \sum_{t=1}^{t_1} n_t$  number of new data symbols to R2.

Before starting the description of the next phase, we introduce some terminology and a lemma that allows us to design the transmission strategy over the next phase.

Let  $\mathcal{S}_{\text{IS}} = \{I_{2[1:n_t+(m_t-N_1)]}(t)\}_{t=1}^{t_1}$  and, as before, an element of it is referred to as the interfering symbol. Further define  $\mathcal{S}_{\text{IS-R2unknown}} = \{I_{2[1:n_t]}(t)\}_{t=1}^{t_1}$  and  $\mathcal{S}_{\text{IS-R2known}} = \{I_{2[n_t+1:n_t+(m_t-N_1)]}(t)\}_{t=1}^{t_1}$  so that  $\mathcal{S}_{\text{IS}} = \mathcal{S}_{\text{IS-R2unknown}} \cup \mathcal{S}_{\text{IS-R2known}}$ . Let  $\{u_{2j}\}_{j=1}^{n'}$  be the i.i.d. complex Gaussian data symbols to be sent by T2 over the next phase and set  $\mathcal{S}_{\text{DS}} = \{u_{2j}\}_{j=1}^{n'}$  and  $\mathcal{S} = \mathcal{S}_{\text{IS}} \cup \mathcal{S}_{\text{DS}}$ .

Consider the following lemma about partitioning of set  $\mathcal{S}$ .

**Lemma 16:** There exists a fixed, deterministic partition  $\mathcal{P}$  of set  $\mathcal{S}$  into  $t_2$  disjoint subsets  $\mathcal{S}(j)$ ,  $j \in [1 : t_2]$ , of cardinality at most  $N_1$  such that each resulting subset  $\mathcal{S}(j)$  satisfies the following properties:

- 1) It does not contain more than  $M_2$  elements of set  $\mathcal{S}_{\text{DS}}$ .
- 2) It does not contain more than  $N_2$  elements of set  $\mathcal{S}_{\text{IS-R2unknown}} \cup \mathcal{S}_{\text{DS}}$ .

*Proof:* This lemma needs to be proved whenever the achievability scheme  $\mathcal{IA}$ -Scheme is to be used and a particular choice for the parameters is made. In the proof of this lemma, the following facts are useful:

$$|\mathcal{S}_{\text{DS}}| = n' = d_2^* - \sum_{t=1}^{t_1} n_t \quad (37)$$

$$|\mathcal{S}_{\text{IS-R2unknown}}| = \sum_{t=1}^{t_1} n_t \quad (38)$$

$$|\mathcal{S}_{\text{IS-R2known}}| = \sum_{t=1}^{t_1} m_t - N_1 t_1 = d_1^* - N_1 t_1. \quad (39)$$

From the proof of Lemma 14 presented Appendix B, it is easy to see that the present lemma holds if the following three inequalities are true: (a)  $M_2 t_2 \geq |\mathcal{S}_{\text{DS}}| = n'$ ; (b)  $N_2 t_2 - |\mathcal{S}_{\text{DS}}| \geq |\mathcal{S}_{\text{IS-R2unknown}}|$ ; and (c)  $N_1 t_2 - |\mathcal{S}_{\text{DS}}| - |\mathcal{S}_{\text{IS-R2unknown}}| \geq |\mathcal{S}_{\text{IS-R2known}}|$ . These inequalities are shown whenever the lemma needs to be proved. ■

Assuming that this lemma is true, we design the next phase of the scheme.

Phase Two: This is the last phase and takes the remaining  $t_2$  time slots. At time  $t \in [t_1 + 1 : t_1 + t_2]$  with  $t' = t - t_1$ , a subset  $\mathcal{S}(t')$ , obtained from the earlier lemma, is chosen. T1 transmits all the interfering symbols belonging to this set (i.e., the elements of set  $\mathcal{S}(t') \cap \mathcal{S}_{\text{IS}}$ ), whereas T2 transmits all data symbols in this set (i.e., the elements of set  $\mathcal{S}(t') \cap \mathcal{S}_{\text{DS}}$ ). As far as T1 is concerned, this is feasible because for no  $t'$  is the cardinality of the subset  $\mathcal{S}(t')$  more than  $N_1$ , which itself is smaller than  $M_1$ . Further, for no  $t' \in [1 : t_2]$  does the subset  $\mathcal{S}(t')$  contain more than  $M_2$  data symbols, and hence, T2 at time  $t$  can transmit the data symbols belonging to the set  $\mathcal{S}(t')$ .

For a  $t \in [t_1 + 1 : t_1 + t_2]$ , T1 and T2 together need to transmit at most  $N_1$  symbols. These symbols are transmitted in a one-to-one fashion (c.f. equation (36)). Therefore, not more than  $N_1$  transmit antennas are in use at any given time during this phase. This implies that R1 via simple channel inversion can recover the required interfering symbols. After recovering the interfering symbols, R1 has one interference-free LC per data symbol and can hence decode the message.

As for R2, the second property of Lemma 16 implies that at most  $N_2$  symbols that are unknown to R2 are transmitted at any given time during this phase. Hence, R2 can subtract the contribution due to the known interfering symbols and then recover the desired data symbols as well as the unknown interfering symbols. Thus, it can also decode all the desired data symbols sent to it over the two phases.

In other words, at the end of  $T$  time slots, the  $i^{\text{th}}$  receiver can decode  $d_i^*$  data symbols which are intended for it (note,  $n' + \sum_{t=1}^{t_1} n_t = d_2^*$ ), provided Lemma 16 holds.

Thus, if the achievability scheme  $\mathcal{IA}$ -Scheme is used with parameters which are such that Lemma 16 holds, then we know that the DoF tuple  $\left(\frac{d_1^*}{T}, \frac{d_2^*}{T}\right)$  is achievable over the  $(M_1, M_2, N_1, N_2)$  MIMO IC.

Now, returning to the achievability of point  $P_{1,3}$  under Case B.II.2, we use achievability scheme  $\mathcal{IA}$ -Scheme with the following choice of parameters:  $T = M_1 - N_2$ ;  $t_1 = N_1 - N_2$ ; which implies  $t_2 = M_1 - N_1$ ;  $d_1^* = M_1(N_1 - N_2)$ ;  $d_2^* = N_2(M_1 - N_2)$ ;  $m_t = M_1 \forall t \in [1 : t_1]$  so that  $\sum_{t=1}^{t_1} m_t = d_1^*$ ;  $n_t$  are deterministic integers such that

$$\sum_{t=1}^{t_1} n_t = \min \{M_2(N_1 - N_2), N_2(M_1 - N_1)\}.$$

It is now sufficient to prove that Lemma 16 holds with these choices of parameters, which is

done in Appendix C. Hence, the DoF pair  $\left(\frac{d_1^*}{T}, \frac{d_1^*}{T}\right)$ , which coincides with  $P_{1,3}$  is achievable. This concludes the proof of Lemma 13. ■

## X. PROOF OF THEOREM 2: CASE B.III

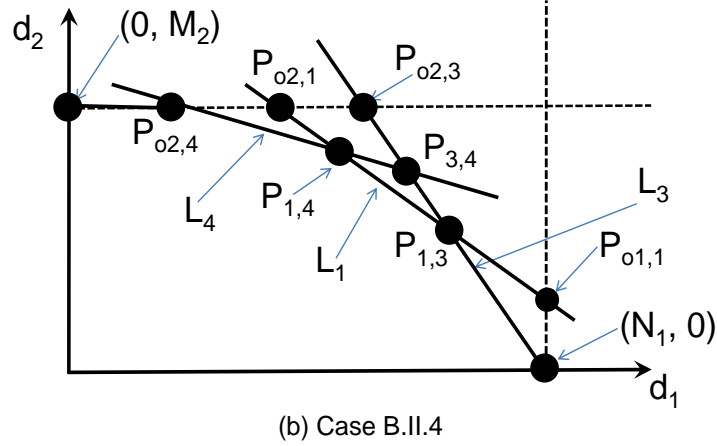
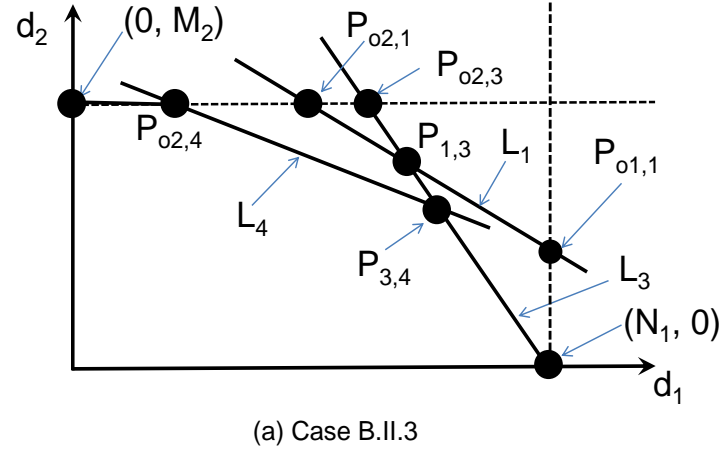


Fig. 6. Two Possible Shapes of the Outer-Bound: Case B.III

Using Lemma 20, the defining inequality of this case is given by

$$M_1 > M'_1 > N_1 > N_2 > M_2 > m.$$

We have the following lemma for this case.

**Lemma 17:** Under Case B.III, bounds  $L_{\{1,3,4\}}$  are active, the outer-bound  $\mathbf{D}_{\text{outer}}^{\text{d-CSI}}$  achievable, and  $\mathbf{D}^{\text{no-CSI}} \subset \mathbf{D}^{\text{d-CSI}} \subset \mathbf{D}^{\text{p-CSI}}$ . Moreover,

1. if  $M_1 \geq N_1 + N_2 - m$ , then bounds  $L_{\{3,4\}}$  are active; and
2. if  $M_1 < N_1 + N_2 - m$ , then bounds  $L_{\{1,3,4\}}$  are active.

*Proof:* We first determine the typical shape of the outer-bound. To this end, note from the previous section that whenever  $M_1 > N_1 > N_2 > M_2$ , the bounds  $L_{\{1,3,4\}}$  are active. Let  $d_i(P_{A,B})$  denote the  $d_i$ -coordinate of point  $P_{A,B}$ . It can be verified that  $d_1(P_{o2,4}) \leq d_1(P_{o2,1}), d_1(P_{o2,3})$  (see Fig. 6). In other words, at  $d_2 = M_2$ , bound  $L_4$  is active. Then it is not difficult to see that if  $d_2(P_{1,3}) \geq d_2(P_{1,4})$ , then bounds  $L_{\{3,4\}}$  are active (see Fig. 6). It turns out that

$$d_2(P_{1,3}) \geq d_2(P_{1,4}) \Leftrightarrow M_1 \geq N_1 + N_2 - m$$

and hence, under Case B.III.1, bounds  $L_{\{3,4\}}$  are active (see Fig. 6(a)). This also implies that under Case B.III.2, bounds  $L_{\{1,3,4\}}$  are active (see Fig. 6(b)).

*A. Case B.III.1: Condition 1 holds and  $M_1 \geq N_1 + N_2 - m$*

From Fig. 6(a), one may deduce that the entire outer-bound can be achieved via time sharing, provided the points  $P_{o2,4}$  and  $P_{3,4}$  can be achieved, which is the topic of the remainder of this subsection.

1) An achievability scheme for  $P_{o2,4}$ : We have

$$P_{o2,4} \equiv \left( \frac{M'_1(N_2 - M_2)}{N_2}, M_2 \right) \text{ with } M'_1 = \min\{M_1, N_1 + N_2 - M_2\} = N_1 + N_2 - M_2.$$

Note that  $(M'_1, M_2, N_1, N_2)$  MIMO IC belongs to Case B.II.2. Therefore, the achievability of point  $P_{o2,1}$ , proved in the context of Case B.II.2, implies the achievability of point  $P_{o2,4}$  for Case B.III.1 (see previous section).

2) An achievability scheme for  $P_{3,4}$ : We have

$$P_{3,4} \equiv \left( N_1 - \frac{N_2^2}{M'_1}, \frac{N_2^2}{M'_1} \right).$$

Note that the point  $P_{3,4}$  remains invariant even if  $M_1$  increases beyond  $N_1 + N_2$ . Moreover, the scheme that we propose here to achieve  $P_{3,4}$  makes use of at most  $N_1 + N_2$  antennas at R1 (even if more are available). Thus, it can be assumed without loss of generality that  $M_1 \leq N_1 + N_2$ .

We first introduce some notation. Let  $\lceil N_1 + N_2 - m \rceil$  ( $\lfloor N_1 + N_2 - m \rfloor$ ) be the smallest (largest) integer greater (smaller) than or equal to  $N_1 + N_2 - m$ . Note that  $N_1 + N_2 - m \leq M_1, N_1 + N_2$  and  $N_1 + N_2 - m \geq N_1 + N_2 - M_2$ .



We will use the scheme  $\mathcal{IA}$ -Scheme with the following choice of parameters:  $T = M'_1$ ;  $t_1 = N_1 - M_2$ , which implies  $t_2 = T - t_1 = N_2$ ;  $d_1^* = N_1 M'_1 - N_2^2$ ; and  $d_2^* = N_2^2$ . Choose  $m_t$  such that  $m_t \in \{ \lceil N_1 + N_2 - m \rceil, \lfloor N_1 + N_2 - m \rfloor \} \forall t \in [1 : t_1]$  and  $\sum_{t=1}^{t_1} m_t = d_1^* = N_1 M'_1 - N_2^2$  and such a choice can be made because  $(N_1 + N_2 - m)t_1 = N_1 M'_1 - N_2^2$ . Next, set  $n_t = N_1 + N_2 - m_t \forall t \in [1 : t_1]$ . This choice ensures that  $m_t \leq M_1$  and  $n_t \leq M_2 \forall t \in [1 : t_1]$ . Note that  $\left(\frac{d_1^*}{T}, \frac{d_2^*}{T}\right)$  is equal to  $P_{3,4}$ .

With this choice of parameters, Lemma 16 is proved below.

*Proof of Lemma 16 for point  $P_{3,4}$  under Case B.III.1:* Since  $n_t = N_1 + N_2 - m_t$ , one can easily compute

$$\sum_{t=1}^{t_1} n_t = N_2(N_2 - M_2),$$

which yields  $n' = N_2 M_2$ . Using these facts and equations (37), (37), and (37), it can be verified that the inequalities (a), (b), and (c) (stated just after equation (37)) hold. ■

Since Lemma 16 holds, we know that the point  $P_{3,4}$  is achievable.

*B. Case B.III.2: Condition 1 holds and  $M_1 < N_1 + N_2 - m$*

From Fig. 6(b), we observe that the entire outer-bound can be achieved via time sharing, provided the points  $P_{o2,4}$ ,  $P_{1,4}$ , and  $P_{1,3}$  can be achieved, which is the topic of the remainder of this subsection.

The achievability of point  $P_{o2,4}$  can be proved in the same manner as done in the previous subsection.

1) An achievability scheme for  $P_{1,3}$ : We have

$$P_{1,3} \equiv \left( \frac{M_1(N_1 - N_2)}{M_1 - N_2}, \frac{N_2(M_1 - N_1)}{M_1 - N_2} \right).$$

Let  $M'_2 = N_1 + N_2 - M_1 < M_2$ . We use the achievability scheme  $\mathcal{IA}$ -Scheme with the following choice of parameters:  $T = M_1 - N_2$ ;  $t_1 = N_1 - N_2$ ;  $t_2 = T = t_1 = (M_1 - N_1)$ ;  $d_1^* = M_1(N_1 - N_2)$ ;  $d_2^* = N_2(M_1 - N_1)$ ;  $m_t = M_1 \forall t \in [1 : t_1]$  so that  $\sum m_t = d_1^*$ ; choose  $n_t$  as deterministic integers such that  $n_t \leq M'_2 \forall t \in [1 : t_1]$  and  $\sum_{t=1}^{t_1} n_t = \min\{M'_2 t_1, d_2^*\}$ . The required lemma, namely, Lemma 16 has been proved in Appendix D-A.

2) An achievability scheme for  $P_{1,4}$ : We have

$$P_{1,4} \equiv \left( \frac{M_1(M'_1 - N_1)}{M'_1 + N_2 - M_1}, N_2 \frac{N_1 + N_2 - M_1}{M'_1 + N_2 - M_1} \right).$$

We use the scheme  $\mathcal{IA}$ -Scheme with the following choice of parameters:  $T = M'_1 + N_2 - M_1$ ;  $t_1 = M'_1 - N_1$ ;  $t_2 = T - t_1 = M'_2$ ;  $d_1^* = M_1(M'_1 - N_1)$ ;  $d_2^* = N_2(N_1 + N_2 - M_1)$ ;  $m_t = M_1$  and  $n_t = M'_2 \forall t \in [1 : t_1]$ . Lemma 16 is proved in Appendix D-B. ■

## XI. CONCLUSION

In this paper, we study the 2-user MIMO IC with delayed CSI. Inner and outer-bounds to its DoF region are obtained and are shown to coincide for all possible values of the 4-tuple  $(M_1, M_2, N_1, N_2)$ . To derive an outer-bound, the property of statistical equivalence of the channel outputs is used, whereas to obtain an inner-bound, interference alignment based achievability schemes, that require only delayed CSIT, are developed.

## APPENDIX A

### LEMMAS USEFUL FOR DETERMINING THE SHAPE OF THE DOF REGION

Consider the following lemma.

**Lemma 18:** For the IC with  $N_1 \geq N_2$  and  $M_1 > N_2$ ,

$$d_1 = \min(M_1, N_1) \Rightarrow d_2 = 0,$$

for the cases of delayed and no CSI.

*Proof:* Since  $\mathbf{D}^{\text{no-CSI}} \subseteq \mathbf{D}^{\text{d-CSI}}$ , it is sufficient to prove the lemma for the case of delayed CSI, which is considered below. If  $N_1 \geq M_2$ , bound  $L_3$  implies that

$$\begin{aligned} d_1 + d_2 &\leq \min \{M_1 + M_2, N_1 + N_2, \max(M_1, N_2), \max(M_2, N_1)\} \\ &= \min \{M_1 + M_2, N_1 + N_2, M_1, N_1\} = \min(M_1, N_1). \end{aligned}$$

Hence, the lemma follows. Now, when  $N_1 < M_2$ , the bounds  $L_1$  and  $L_3$  are given by

$$L_1 \equiv \frac{d_1}{\min(M_1, N_1 + N_2)} + \frac{d_2}{N_2} \leq 1 \text{ and } L_2 \equiv \frac{d_1}{N_1} + \frac{d_2}{\min(N_1 + N_2, M_2)} \leq 1.$$

If  $N_1 \leq M_1$ ,  $L_2$  implies that  $d_2 = 0$  whenever  $d_1 = \min(M_1, N_1) = N_1$ ; whereas  $N_1 > M_1$ , the same is implied by  $L_1$ . ■

**Lemma 19:** Consider the MIMO IC with delayed CSI for which  $N_1 \geq N_2$  and  $M_1 > N_2$ . Under Case A, where  $M_2 \geq N_2$ , we have

$$d_2 = \min(M_2, N_2) \Rightarrow d_1 = 0,$$

which is not true with Case B, where  $M_2 < N_2$ .

*Proof:* Whenever  $N_1 \geq N_2$  and  $M_1 > N_2$ , bound  $L_1$  is given by

$$L_1 \equiv \frac{d_1}{\min(M_1, N_1 + N_2)} + \frac{d_2}{N_2} \leq 1.$$

Therefore, under Case A, if  $d_2 = \min(M_2, N_2) = N_2$ ,  $d_1$  can not be more than zero, which proves the first part of the lemma. Under Case B, when  $d_2 = \min(M_2, N_2) = M_2$ , one can achieve  $d_1 = N_2 - M_2$ , since  $N_1, M_1 \geq N_2$ , even without CSI [16]. This proves the second part of the lemma. ■

**Lemma 20:** Condition 1 can equivalently be stated as

$$M_1 > M'_1 > N_1 > N_2 > M_2 > m,$$

where  $M'_1 = \min(M_1, N_1 + N_2 - M_2)$  and  $m = N_2 \frac{M'_1 - N_1}{M'_1 - N_2}$ .

*Proof:* Under Condition 1,  $M'_1 = N_1 + N_2 - M_2$ , which implies that  $M'_1 - N_1 = N_2 - M_2$  and  $M'_1 - N_2 = N_1 - M_2$ . Hence, Condition 1 implies the inequality stated in this lemma. Further, if the inequality stated in the lemma is true, then  $M_1 > N_1 + N_2 - M_2$ , which then implies Condition 1. ■

## APPENDIX B

### PROOF OF LEMMA 14

As a first step towards obtaining the desired partition of the set  $\mathcal{S}$ , we partition the set  $\mathcal{S}_{\text{R2-unknown}}$  into  $t_2$  subsets  $\{\mathcal{S}_{\text{R2-unknown}}(j)\}_{j=1}^{t_2}$  of cardinality at most  $N_2 - M_2$ . This is feasible as per the following lemma.

**Lemma 21:**  $|\mathcal{S}_{\text{R2-unknown}}| = M_2(N_1 - M_2) \leq (N_2 - M_2)t_2$ .

*Proof:* The fact that  $|\mathcal{S}_{\text{R2-unknown}}| = M_2(N_1 - M_2)$  can be easily verified by noting that  $t_1 = N_1 - M_2$ . Suppose the inequality stated in the lemma is not true, i.e., the inequality  $M_2(N_1 - M_2) > (N_2 - M_2)(M'_1 + M_2 - N_2)$  holds (recall  $t_2 = (M'_1 + M_2 - N_2)$ ). This yields us

$$\begin{aligned} M_2(N_1 - M_2) > (N_2 - M_2)(M'_1 + M_2 - N_2) &\Leftrightarrow \frac{N_1 - M_2}{M'_1 + M_2 - N_1} > \frac{N_2 - M_2}{M_2} \\ &\Leftrightarrow \frac{M'_1}{M'_1 + M_2 - N_1} > \frac{N_2}{M_2} \Leftrightarrow \frac{M_2}{N_2} > \frac{M_2 + (M'_1 - N_1)}{N_2 + (M'_1 - N_2)} \Leftrightarrow M_2 > N_2 \frac{M'_1 - N_1}{M'_1 - N_2} = m, \end{aligned}$$

which contradicts the defining inequality of Case B.II.1. Hence, the inequality stated in lemma holds. ■

Consider now the following algorithm that yields the required partition  $\mathcal{P}$ .

- 1) Set  $j = 1$  and  $x = |\mathcal{S}_{\text{R2-unknown}}(j)|$ .
- 2) Compute  $x' = (N_1 - M_2) - x$ . Pick  $x'$  distinct elements of set  $\mathcal{S}_{\text{R2-known}}$  and let  $s$  be the set of chosen elements. Set  $\mathcal{S}(j) = \mathcal{S}_{\text{R2-unknown}}(j) \cup s$  and  $\mathcal{S}_{\text{R2-known}} = \mathcal{S}_{\text{R2-known}} \setminus s$ .
- 3) If  $j = t_2$ , stop; else increment  $j$  by 1 and go to Step 1 above.

Note here that a given element of set  $\mathcal{S}_{\text{R2-known}}$  can belong to a subset  $\{\mathcal{S}(j)\}_j$  for at most one  $j$  because at every step distinct elements are chosen and once an element is chosen it is removed from further consideration. Hence, the resulting subsets are disjoint. Further, the algorithm can

accommodate

$$\begin{aligned}
\sum_{j=1}^{t_2} \left\{ (N_1 - M_2) - |\mathcal{S}_{\text{R2-unknown}}(j)| \right\} &= t_2(N_1 - M_2) - |\mathcal{S}_{\text{R2-unknown}}| \\
&= t_2(N_1 - M_2) - M_2(N_1 - M_2) = (t_2 - M_2)(M'_1 - N_1) \\
&= (M'_1 - N_2)(M'_1 - N_1)
\end{aligned}$$

elements of set  $\mathcal{S}_{\text{R2-known}}$ . Since  $|\mathcal{S}_{\text{R2-known}}| = (N_1 - M_2)(M'_1 - N_1)$  and since the subsets  $\{\mathcal{S}(j)\}_j$  have been shown to be disjoint, and every element of set  $\mathcal{S}_{\text{R2-known}}$  belongs to  $\{\mathcal{S}(j)\}$  for exactly one  $j \in [1 : t_2]$ . Hence, the algorithm yields a partition with the required properties.

## APPENDIX C

### PROOF OF LEMMA 16 FOR POINT $P_{1,3}$ UNDER CASE B.II.2

We consider two cases separately: first, when  $M_2(N_1 - N_2) \geq N_2(M_1 - N_1)$ , and second when  $M_2(N_1 - N_2) < N_2(M_1 - N_1)$ .

A.  $M_2(N_1 - N_2) \geq N_2(M_1 - N_1)$ :

In this case,  $\sum_{t=1}^{t_1} n_t = N_2(M_1 - N_1)$ , i.e.,  $n' = 0$  or no data symbols needs to be sent to R2 over Phase Two. Thus, the first property stated in the lemma is satisfied trivially. Now, we have

$$\begin{aligned}
|\mathcal{S}_{\text{IS-R2unknown}}| &= \sum_{t=1}^{t_1} n_t = N_2(M_1 - N_1) = N_2 t_2, \\
|\mathcal{S}_{\text{IS-R2known}}| &= (M_1 - N_1)(N_1 - N_2), \text{ and } |\mathcal{S}_{\text{DS}}| = 0.
\end{aligned}$$

The second property stated in the lemma dictates that at most  $N_2 t_2$  elements of the set  $\mathcal{S}_{\text{IS-R2unknown}}$  can be accommodated into subsets  $\{\mathcal{S}(j)\}_{j=1}^{t_2}$  (given that  $|\mathcal{S}_{\text{DS}}| = 0$ ). Since  $|\mathcal{S}_{\text{IS-R2unknown}}| = N_2 t_2$ , the elements of the set  $\mathcal{S}_{\text{IS-R2unknown}}$  can be distributed into subsets  $\{\mathcal{S}(j)\}_j$  such that every element of it belongs to  $\mathcal{S}(j)$  for exactly one  $j$ . After this step, the subsets  $\{\mathcal{S}(j)\}_j$ , each of which is of cardinality at most  $N_1$ , can accommodate at most  $N_1 t_2 - N_2 t_2 = (N_1 - N_2) t_2$  elements of set  $\mathcal{S}_{\text{IS-R2known}}$ , which equals the cardinality of this set itself. Hence, the partition  $\mathcal{P}$  with the required properties can be obtained. See also the algorithm presented in Appendix B, which can be suitably modified to apply here.

B.  $M_2(N_1 - N_2) < N_2(M_1 - N_1)$ :

Here,  $n' > 0$ , i.e., T2 needs to transmit new data symbols during Phase Two. Let us first focus on the first property, which says that at most  $M_2 t_2$  elements of set  $\mathcal{S}_{\text{DS}}$  can be accommodated in the subsets resulting  $\{\mathcal{S}(j)\}_j$ . But,

$$\begin{aligned} M_2 t_2 = M_2(M_1 - N_1) &= M_2(M_1 - N_2) - (N_1 - N_2)M_2 \\ &> N_2(M_1 - N_1) - (N_1 - N_2)M_2 = |\mathcal{S}_{\text{DS}}|, \end{aligned}$$

where the inequality follows from the fact that under Case B.II.2,  $M_2 > N_2 \frac{M_1 - N_1}{M_1 - N_2}$ . Hence, as a first step, the elements of set  $\mathcal{S}_{\text{DS}}$  are distributed into subsets  $\{\mathcal{S}(j)\}_{j=1}^{t_2}$  such that no subset contains more than  $M_2$  elements and every element belongs to exactly one subset.

Let us now focus on the second property stated in the lemma, which says that at most  $N_2 t_2$  elements of the set  $\mathcal{S}_{\text{IS-R2unknown}} \cup \mathcal{S}_{\text{DS}}$  can be accommodated into subsets  $\{\mathcal{S}(j)\}_{j=1}^{t_2}$ . After having performed the first step, the subsets  $\{\mathcal{S}(j)\}$  can contain a total of

$$\begin{aligned} N_2 t_2 - |\mathcal{S}_{\text{DS}}| &= N_2(M_1 - N_1) - \{N_2(M_1 - N_1) - (N_1 - N_2)M_2\} \\ &= M_2(N_1 - N_2), \end{aligned}$$

elements of  $\mathcal{S}_{\text{IS-R2unknown}}$ , which equals its cardinality. Having performed the first step above, the elements of the set  $\mathcal{S}_{\text{IS-R2unknown}}$  can be distributed into subsets  $\{\mathcal{S}(j)\}$  such that the two properties stated in the lemma hold and every element of  $\mathcal{S}_{\text{IS-R2unknown}} \cup \mathcal{S}_{\text{DS}}$  belongs to exactly one subset.

At the final step, the elements of  $\mathcal{S}_{\text{IS-R2known}}$  are distributed into subsets  $\{\mathcal{S}(j)\}$  such that every element belongs to exactly one subset. This can be done because the resulting subsets can accommodate at most  $N_1 t_2 - \{|\mathcal{S}_{\text{DS}}| + |\mathcal{S}_{\text{IS-R2unknown}}|\} = N_1 t_2 - N_2 t_2 = (N_1 - N_2)t_2 = (M_1 - N_1)(N_1 - N - 2) = |\mathcal{S}_{\text{IS-R2known}}|$  elements of the set  $\mathcal{S}_{\text{IS-R2known}}$ , i.e., all the elements of  $\mathcal{S}_{\text{IS-R2known}}$  can be accommodated.

## APPENDIX D

### PROOF OF LEMMA 16 FOR CASE B.III.2

A. For Point  $P_{1,3}$

If  $M'_2(N_1 - N_2) \geq N_2(M_1 - N_1)$  (i.e.,  $n' = 0$ ), the proof given in Appendix C-A holds without any change. Consider thus the case of  $n' > 0$ . First, we prove the inequality (a), i.e.,  $M_2 t_2 \geq n'$ .

Suppose the inequality does not hold. Then we get

$$\begin{aligned}
M_2(M_1 - N_1) &< n' = N_2(M_1 - N_1) - (N_1 + N_2 - M_1)(M_1 - N_1) \\
&\Rightarrow (N_2 - M_2)(M_1 - N_1) > (N_1 + N_2 - M_1)(N_1 - N_2) \\
&\Rightarrow N_2(N_1 - N_2) < (M_1 - N_1)(N_1 - M_2) \\
&\Rightarrow N_2(N_1 - N_2) < (N_1 - M_2)(N_2 - N_2 \frac{N_2 - M_2}{N_1 - M_2}) \\
&\Rightarrow N_2(N_1 - N_2) < N_2(N_1 - N_2),
\end{aligned} \tag{40}$$

where the inequality (40) holds since under Case B.III.2, we have  $M_1 < N_1 + N_2 - m$ . The last inequality yields us a contradiction, which implies that the desired inequality (a) is true. The remaining two inequalities, namely, (b) and (c) can be easily proved and the details have been omitted.

#### B. For Point $P_{1,4}$

We can compute  $|\mathcal{S}_{\text{DS}}| = N_2 M'_2 - M'_2(M'_1 - N_1) = M_2 M'_2 = M_2 t_2$ , which implies nequality (a). Now,  $|\mathcal{S}_{\text{IS-R2unknown}}| = M'_2 t_1 = M'_2(M'_1 - N_1) = M'_2(N_2 - M_2) = N_2 t_2 - |\mathcal{S}_{\text{DS}}|$ , which implies inequality (b). To prove the last inequality, we need to show that  $|\mathcal{S}_{\text{IS-R2known}}| = (N_2 - M'_2)(M'_1 - N_1) \leq M'_2(N_1 - N_2)$ . Suppose assume the contrary. Then we have

$$\begin{aligned}
\frac{N_1 - N_2}{M_1 - N_1} &< \frac{N_2 - M_2}{N_1 + N_2 - M_1} \Rightarrow \frac{M_1 - N_2}{M_1 - N_1} < \frac{N_1 + 2N_2 - M_1 - M_2}{N_1 + N_2 - M_1} \\
&\Rightarrow \frac{N_1 + N_2 - M_1}{M_1 - N_1} < \frac{N_1 + 2N_2 - M_1 - M_2}{M_1 - N_2} \Rightarrow \frac{N_1 + N_2 - M_2}{M_1 - N_2} > \frac{N_2}{M_1 - N_1} \\
&\Rightarrow \frac{M_1 - N_1}{M_1 - N_2} > \frac{N_2}{N_1 + N_2 - M_2} \Rightarrow M_1 > N_1 + N_2 \frac{N_1 - N_2}{N_1 - M_2} \\
&\Rightarrow M_1 > N_1 + N_2 \frac{N_1 - M_2 - [N_2 - M_2]}{N_1 - M_2} \Rightarrow M_1 > N_1 + N_2 - m,
\end{aligned}$$

which is a contradiction since under Case B.III.2, we have  $M_1 < N_1 + N_2 - m$ . Hence, inequality (c) holds.

## REFERENCES

- [1] A. B. Carleial, "A case where interference does not reduce the capacity," *IEEE Trans. on Inform. Th.*, vol. 21, pp. 569–570, Sep, 1975.
- [2] R. Benzal, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. on Inform. Th.*, vol. 25, pp. 228–231, Mar, 1979.

- [3] T. S. Han and K. Kobayashi, "A new achievable region for the interference channel," *IEEE Trans. on Inform. Th.*, vol. 27, pp. 49–60, Jan, 1981.
- [4] H. Sato, "The capacity of Gaussian interference channel under strong interference," *IEEE Trans. on Inform. Th.*, vol. 27, pp. 786–788, Nov, 1981.
- [5] A. A. E. Gamal and M. H. M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. on Inform. Th.*, vol. 28, pp. 343–346, Mar, 1982.
- [6] V. S. Annapureddy and V. V. Veeravalli, "Sum capacity of the Gaussian interference channels in the low interference regime," *IEEE Trans. on Inform. Th.*, vol. 55, pp. 3032 – 3050, Jul. 2009.
- [7] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Trans. on Inform. Th.*, vol. 55, pp. 620–643, Feb, 2009.
- [8] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. on Inform. Th.*, vol. 55, pp. 689–699, Feb, 2009.
- [9] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
- [10] E. Telatar and D. N. C. Tse, "Bounds on the capacity region of a class of interference channels," in *Proc. IEEE Int. Symp. on Inform. Th.*, Jun, 2007.
- [11] S. Karmakar and M. K. Varanasi, "Capacity within constant gap of the MIMO interference channel," 2011, Available Online: Arxiv.
- [12] S. A. Jafar and M. J. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.
- [13] C. Huang and S. A. Jafar, "Degrees of freedom of the MIMO interference channel with cooperation and cognition," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 4211–4220, Sep. 2009.
- [14] C. Huang, S. A. Jafar, S. Shamai, and S. Vishwanath, "On degrees of freedom region of MIMO networks without CSIT," Sep. 2009, Available Online: <http://arxiv.org/pdf/0909.4017>.
- [15] Y. Zhu and D. Guo, "The degrees of freedom of MIMO interference channels without state information at transmitters," Aug. 2010, [Online.] Available: <http://arxiv.org/abs/1008.5196>.
- [16] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," Sep. 2009, Available Online: <http://arxiv.org/abs/0909.5424>.
- [17] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [18] C. S. Vaze and M. K. Varanasi, "CSI feedback scaling rate vs multiplexing gain tradeoff for DPC-based transmission in the Gaussian MIMO broadcast channel," in *IEEE Inter. Symp. Inform. Theory*, Jun. 2010.
- [19] R. T. Krishnamachari and M. K. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," Nov. 2009, Available: <http://arxiv.org/abs/0911.5509>.
- [20] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," Oct. 2010, Available: <http://arxiv.org/abs/1010.1499>.
- [21] H. Maleki, S. A. Jafar, and S. Shamai, "Retrospective interference alignment," Sep. 2010, Available: [http://arxiv.org/PS\\_cache/arxiv/pdf/1009/1009.3593v1.pdf](http://arxiv.org/PS_cache/arxiv/pdf/1009/1009.3593v1.pdf).
- [22] C. S. Vaze and M. K. Varanasi, "Degrees of freedom region for the two-user MIMO broadcast channel with delayed CSI," Available: <http://arxiv.org/abs/1101.0306>, Dec. 2010.



- [23] S. A. Jafar, "Exploiting channel correlations - simple interference alignment schemes with no CSIT," Oct. 2009, Available Online: <http://arxiv.org/pdf/0910.0555>.
- [24] A. E. Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 379–381, Apr. 1978.
- [25] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.
- [26] C. S. Vaze and M. K. Varanasi, "The degrees of freedom region of the MIMO cognitive interference channel with no CSIT," in *IEEE Inter. Symp. Inform. Theory*, Austin, USA, June 2010.
- [27] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Euro. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [28] Y. Zhu and D. Guo, "Isotropic MIMO interference channels without CSIT: The loss of degrees of freedom," in *Forty-Seventh Annual Allerton Conference*, UIUC, IL, USA, Sep./Oct. 2009.
- [29] T. Cover and J. Thomas, *Elements of Inform. Theory*. John Wiley and Sons, Inc., 1991.
- [30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1985.